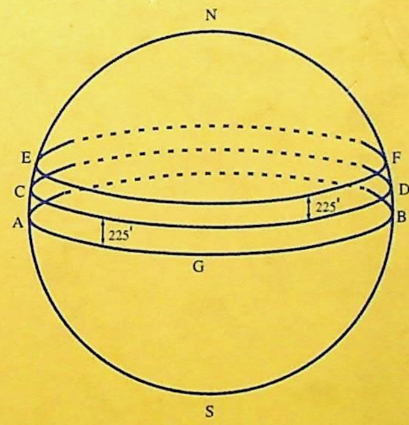
Culture and History of
Mathematics 5

Studies in the History of Indian Mathematics

C. S. Seshadri
Editor



									१
							१	५	१५
					१	४	१०	२०	३६
			१	३	६	१०	१८	३३	६१
	१	२	३	४	७	१२	२१	३४	५४
१	१	१	२	३	५	७	१०	१४	२१
१	२	३	४	५	६	७	८	९	१०

HINDUSTAN
BOOK AGENCY

CULTURE AND HISTORY OF MATHEMATICS 5

Studies in the History of Indian Mathematics

CULTURE AND HISTORY OF MATHEMATICS

Editor
Rajendra Bhatia, Indian Statistical Institute, New Delhi.

Already Published Volumes
Bruce C. Berndt and Robert A. Rankin (editors): Ramanujan: Essays and Surveys
Richard S. Palais: Seminar on the Atiyah-Singer Index Theorem
Gerard G. Emch, R. Sridharan and M. D. Srinivas (editors): Contributions to the History of Indian Mathematics
K. V. Sarma: Gaṇita-Yukti-Bhāṣā (Rationales in Mathematical Astronomy) of Jyeṣṭhadeva

Studies in the History of Indian Mathematics

Editor

C. S. Seshadri

HINDUSTAN
BOOK AGENCY

Published in India by
Hindustan Book Agency (India)
P 19 Green Park Extension
New Delhi 110 016
India

email: info@hindbook.com
http://www.hindbook.com

ISBN 978-93-80250-06-9

Contents

Editorial Note

It gives me immense satisfaction that we could assemble a group of distinguished scholars at this Institute to conduct a seminar on the history of mathematics in India, leading to the publication of this volume. This whole endeavour owes very much to the initiative and commitment of David Mumford. I am very glad that this provided also an occasion for having amidst us the renowned indologist Frits Staal.

It is fitting that this volume is dedicated to the memory of the two outstanding scholars David Pingree and K.V. Sarma who have made pioneering contributions to this field.

It is my pleasant duty to thank all those who have contributed to this volume, as well as to the one invisible contributor Jayant Shah (Northeastern University, Boston) for his critical reading of all the material in this volume.

C.S. Seshadri
Director
Chennai Mathematical Institute

Preface

It was a great privilege to participate in the Seminar *Topics in the History of Indian and Western Mathematics* which took place at the Chennai Mathematical Institute in January and February 2008. I am a newcomer in this area and cannot read Sanskrit, but I had the opportunity to listen and learn from Sanskrit scholars, mostly Indian but one Western (Staal) and absorb their various perspectives. It has been very exciting to me over the last five years to learn something of the distinctive Indian approach to mathematics, from its beginnings in Vedic times to its wonderful achievements in algebra and the Calculus, just prior to the waves of Western invasions starting in the 16^{th} century.

Today, there is a major resurgence of scholarship in Indian mathematics and astronomy both in India and the West, led, on the one hand, by a widespread wave of renewed interest in India in Sanskrit studies and on the other hand, by the school created by my late colleague David Pingree. Until recently, nearly all articles and books on the History of Indian Mathematics and Astronomy were nearly impossible to find in the West. But now, new translations and critical editions with commentary of many of the extant manuscripts are being published and widely distributed in the West as well as in India.

This volume contains analyses of many of the most important topics in Indian Mathematics and Astronomy, taken from talks at this seminar. Let me sketch some of the ideas from each chapter, highlighting topics which seemed to me especially significant. I will follow a roughly chronological order.

Dani's chapter deals with the oldest extant Indian mathematical works, the *Śulbasūtras*, or "Rules of the Cord", which date from as early as c.800 BCE. These are manuals on the geometry needed for erecting the fire altars central to Vedic ritual. Dani describes in detail their techniques of geometric algebra, their construction for a square whose area is the sum of those of two given squares or whose area is the same as that of a given rectangle. These anticipate many of the constructions which appear later in Book II of Euclid's *Elements*, with knotted cords replacing the Greek's use of straight edge and compass. In particular, "Pythagoras's" theorem is described here well before Pythagoras. (But note that it also occurs as early as c.1800 BCE on Babylonian tablets.) Dani then goes on to describe their approximate constructions of circles whose area is that of a given square and vice versa. His last section deals with the rational approximation given for the square root of two given in three of the four *Śulbasūtras*. Like the Babylonian approximation, it is accurate to roughly one part in a million. Dani describes a striking geometric method, first proposed by Datta, by which this approximation might have been found. But he also proposes that, since they were using very long ropes to actually lay out fire altars, it is not impossible that they could have carried out a continued fraction-like algorithm using two ropes and repeatedly taking multiples of the shorter away from the longer. The idea that such algorithms have a very long history connects to Dutta's article.

Staal—the other Westerner besides myself in this seminar—considers the origins of the number zero and proposes a link of the mathematical zero with the linguistic zero markers which were invented by the great grammarian Pāṇini. The foundations of modern computer science as well as linguistics go back to Pāṇini (c.400 BCE). Pāṇini invented formal grammar with abstract variables for various parts of an utterance and recursive rewrite rules. Sanskrit grammar was well-known to all the mathematician–astronomers and is a plausible source for many ideas which were later developed in more mathematical ways. Staal focuses on the Sanskrit word *lopa*, "something that does not appear", as a precursor to the idea of zero. As he has written in other works, Sanskrit grammar appears to grow out of the precisely formalized Vedic rituals. Astonishingly, the Vedas dealt with possible enactments of a ritual in which one participant fails to utter some required sentence. The verbal root *lup-* of *lopa* is used to describe this failure: a zero in the ritual. Thus the original zero could be the priest's lapse.

Raja Sridharan, R. Sridharan and M. D. Srinivas's chapter concerns another area of Indian mathematics: the combinatorics which was inspired first by the study of Sanskrit prosody and later by the study of musical patterns, both tonal and rhythmic. Pāṇini was followed by Piṅgaḷa (c.300 BCE) who studied Sanskrit prosody. Sanskrit was traditionally written in verse and memorized. Each line had a characteristic pattern of short and long syllables. Piṅgaḷa devised a way to order all patterns of short and long syllables in a line with n syllables, and, using this, (i) to compute the number describing each pattern, (ii) to reconstruct the pattern from its number and finally (iii) to compute the number of patterns with a fixed number of short, respectively, long syllables. The first and second use the binary number system and the third involves calculating the binomial coefficients. This beautiful foundation led to much further combinatorial work which the authors survey. For instance the Fibonacci numbers arose when they asked how many sequences have a given total length if long syllables are given length 2 and short length 1. Virahāṅka discovered the Fibonacci numbers and their recursion relation in the 7th century (well before Fibonacci!). Recursion seems to be a general theme which runs through much of Indian mathematics. Most of their chapter concerns the generalization to musical phrases and to musical rhythms, where the combinatorics gets more complex and more interesting. This story is beautifully described by the authors.

My own chapter concerns the introduction of negative numbers both in India, in China, in Greece and in modern Europe. The full arithmetic of negative numbers appears in Brahmagupta's *Brāhma-sphuṭa-siddhānta* in the 6th century and presumably arose much earlier, maybe even in the accounting practices described in Kauṭilya's *Arthaśāstra*. In contrast, the first place where this is correctly described without hesitation[1] in modern Europe is in Wallis's *Treatise on Algebra* in 1685.

[1]Other Westerners wondered whether $(-1) \cdot (-1)$ might be taken to equal -1.

This seems to me a stunning example of how far two mathematical traditions can diverge—though converging in the end.

Dutta's contribution concerns Indian arithmetic and algebra. More specifically, it concerns three algorithms. The first, called *kuṭṭaka* (the pulverizer) constructs, for all positive integers a, b, positive integers x, y such that $ax - by$ equals plus or minus the greatest common divisor c of a and b. The first half of the algorithm is the same as the Euclidean algorithm, successive subtraction of the less from the greater, and the second half works backwards to find x and y—just as we do today. This second step was not taken by the Greeks. Kuṭṭaka appears first in rather cryptic form in the *Āryabhaṭīya* (499 CE) and soon after more explicitly in the works of Bhāskara I and Brahmagupta. They use it to construct all positive integral solutions of linear equations $ax - by = d,\ a, b, d$ of positive integers. This had long been a concern because of the ancient Vedic method for making sense of the relative periods of the day, the lunar month, the year and the periods in planetary motion: that at the beginning of the present Kaliyuga, all the planets, the sun and moon were all lined up in one spectacular conjunction.

The second and third algorithms are concerned with "Pell's" equation $x^2 - Dy^2 = 1$ which quite plausibly arose in the search for good rational approximations x/y to $\sqrt{D}$. The second algorithm, due to Brahmagupta and called the *Bhāvāna*, is equivalent, in modern terms, to the law for multiplication of the algebraic integers $x + y\sqrt{D}$. As they put it, if $x^2 - Dy^2 = m$, $u^2 - Dv^2 = n$, then $s = xu + Dyv$, $t = xv + yu$ solves $s^2 - Dt^2 = mn$. They now played with solutions of the equations $x^2 - Dy^2 = m$ for various m and several centuries later, they did indeed find an algorithm which always finds solutions with $m = 1$, the *Cakravāḷa*. Thus they had a complete theory of Pell's equation, modulo one point—a *proof* that this worked. It is interesting that the standard proofs in modern texts are non-constructive whereas the Indian mathematicians focused instead on seeking on constructive methods and never studied non-constructive arguments. Their idea of mathematics was closer to that of applied mathematicians and computer scientists than that of pure mathematicians. Indeed the *siddhāntas* (treatises) where these algorithms were written down were manuals for actually calculating astronomical events.

Ramasubramanian and Srinivas take up the story of Indian work on Calculus. I personally feel this is a story which deserves to be much more widely known in the West. Their chapter outlines the millennium long history of these discoveries, culminating in the complete analysis of the basic calculus of polynomial and trigonometric functions, their integrals and derivatives and power series for sine, cosine and arctangent—and, of course, applications of these to astronomy. The path they took to this is quite disjoint from the path that was taken, first by Archimedes and later by Newton and Leibniz. It is extremely unlikely that any of Archimedes' work on, e.g. the Riemann sum for the integral of sine, made its way to India. Instead the Indian work seems to have taken off with their discovery of the second

order finite difference equation for sine in the 5th century CE or earlier (the discrete analog of the result that sine solves the harmonic equation $y'' + y = 0$.) This appears cryptically in the *Aryabhaṭīya*, and clearly soon after, as is well described in Section 5 of Ramasubramanian and Srinivas's chapter. Also, early was an interest in summing powers of natural numbers which is described in Section 4. This early work revolves around finite differences and the corresponding sums, much like Leibniz's starting point for calculus. But the application to astronomy made it clear by the 10th century that one needed to calculate the "instantaneous velocity" as well as its finite difference approximation (see Section 6). In the 12th century, Bhāskara II used these ideas to rediscover Archimedes' derivation of the formula for the area and volume of a sphere. This derivation reduces the problem to computing the integral of sine. But, interestingly, as an applied mathematician, he completes the proof by numerically summing his sine table—even though he knows quite well that cosine differences are sines and could have done it this way. The numerical method was apparently more convincing!

The crowning achievements of this work are due to a major genius who is nearly unknown in the West: *Mādhava*, who lived in a village in Kerala in the 14th century. Only a small fragment of his work survives, but, fortunately, his and his school's work was written up in an unusual expository form in the local language, Malayalam, by Jyeṣṭhadeva in the 16th century—still over a century before Newton and Leibniz did their work. This book, the *Yukti-bhāṣā* has only now been translated into English by K. V. Śarma with commentary by the authors of this chapter and M. S. Sriram. The first volume was released with some ceremony during our seminar. The work of the Kerala school is described in Part 2 of the present chapter. Let me only mention that in addition to deriving the power series for sine and cosine and the 'Gregory' series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

their numerical bent appears again in finding a series of ways to estimate the remainder so that this becomes a practical tool for calculating π. This led them to much more rapidly converging series such as:

$$\frac{\pi}{4} = \frac{3}{4} + \frac{1}{3^3 - 3} - \frac{1}{5^3 - 5} + \frac{1}{7^3 - 7} - \cdots$$

$$\frac{\pi}{16} = \frac{1}{1^5 + 4 \cdot 1} - \frac{1}{3^5 + 4 \cdot 3} + \frac{1}{5^5 + 4 \cdot 5} - \cdots$$

Divakaran's article proposes the compelling thesis that recursion is the central theme and technique that runs through all the Indian work in mathematics. He traces this to Pāṇini's grammatical rules and even earlier to Vedic ritual. He points out that the whole idea of decimal place-value notation can be seen as a way to describe integers recursively. Recursive generation of larger and larger numbers leads to a

clear conception of infinitely large and its inverse, to a conception of the infinitely small and of the limiting process. The bulk of his article, however, focuses on the extensive use of recursion in the Kerala work on Calculus. Like Newton, he sees the introduction of power series as an algebraic analog of decimal expansions (x^n being the analog 10^n) and he discusses at length the power series expansion

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots$$

found in the *Yukti-bhāṣā*. He then describes the huge step which one finds in the *Yukti-bhāṣā*: recursive proofs based on induction: they derive the integral of x^n using induction on n. Finally, he discusses the derivation of the power series for sine and cosine in the *Yukti-bhāṣā* and notes that it is derived by first converting the known difference equations for sine in terms of cosine and cosine in terms of sine into a summation form. This is the exact analog of converting differential equations into integral equations. Jyeṣṭhadeva, starting with a crude approximation, then recursively back-substitutes each approximation into the summation form to get a better one. Passing to the limit, he gets the power series. This back substitution is exactly what we do today to solve Fredholm integral equations.

Finally, M. S. Sriram's Chapter addresses the planetary models in Indian Astronomy. As mentioned above, mathematics was usually studied together with astronomy and the two subjects advanced hand in hand. His article concentrates on their last model which is due to Nīlakaṇṭha c.1500 CE, who worked at roughly the same time as Copernicus in the West. His model is explained at length in the *Gaṇita-yukti-bhāṣā* and is described with modern formulas in Sriram's chapter. The most interesting issue is here is the historical movement from geocentric models to heliocentric models. In fact, Nīlakaṇṭha's model is 'essentially' heliocentric. In his model, one starts with the mean sun moving in the ecliptic circle. Then for each planet, one takes a plane centered at the mean sun but inclined to it to different amounts for each planet and intersecting it at the appropriate nodes. And in each of these planes one places the corresponding planet on the Ptolemaic approximation to the ellipse: an eccentric circle. Once one overcomes the dense tangle of compound Sanskrit words, a very modern model shines through.

These articles together cover a substantial portion of the exciting History of Indian Mathematics and Astronomy. Only a fraction of this has become generally known to mathematicians in the West. Too many people still think that mathematics was born in Greece and more or less slumbered until the Renaissance. Therefore, I hope that this book may serve as a way of bringing to the international mathematical community a deeper knowledge of the riches in Indian Mathematics. To scholars, however, there is another message: there is much work still to be done to study, edit and translate the many ancient manuscripts still surviving in libraries all over India. One hopes for a deeper and broader picture of the more than two millennium long history when every one of these has been looked at and analyzed.

David Mumford



David Pingree, 1933–2005

Kim Plofker

David Edwin Pingree (2 January 1933–11 November 2005) employed the fifty years of his scholarly career investigating the development of mathematics, astronomy and the related exact sciences from ancient Mesopotamia to early modern Europe and India. He published editions, translations and studies of source texts in Akkadian cuneiform, Greek, Latin, Sanskrit, Arabic and Persian, on subjects ranging from infinite series and interpolation techniques to astral magic and iconography in astrological texts. He was professionally affiliated with Harvard University as an undergraduate (B.A. in Classics and Sanskrit, 1954), graduate student (Ph.D. in Sanskrit and Indian Studies, 1960), and Junior Fellow (to 1963); the University of Chicago as a faculty member in the Oriental Institute and Departments of History, South Asian Languages, and Near Eastern Languages (1963–1971); and Brown University as a professor in the Departments of Classics and the History of Mathematics (1971–2005). Over the course of his immensely productive career he received many honors, including a Fulbright Scholarship, a Guggenheim Fellowship, and membership in several learned societies, among them the American Academy of Arts and Sciences, the American Philosophical Society, and the Institute for Advanced Study at Princeton University. Pingree was awarded the title of "Abhinavavarāhamihira" by the government of Uttar Pradesh in 1979, and in 1981 was one of the first recipients of the MacArthur Fellowship (popularly nicknamed the "Genius Grant"), together with co-honorees including the philosopher Richard Rorty, the paleontologist Stephen Jay Gould, and the computer scientist Stephen Wolfram. His total scholarly output comprised several dozen monographs and several hundred research articles, reviews, encyclopedia entries, and other works.

In his research on the history of science, David Pingree was first and foremost what might be termed a "transmissionist": he was primarily interested in what he called the "kinematics" of scientific development, the ways that scientific ideas were passed from one culture to another, and how they were transformed in the process. The first awakening of his interest in the history of science as a field of research was due, as he described it, to just such a question of transmission from the medieval Indian exact sciences to their Greek counterpart. As a Fulbright scholar in 1955, reading a Byzantine Greek manuscript on astrology in the Vatican Library, he noticed marginal notes by a commentator that included technical terms transliterated from Sanskrit. This discovery eventually inspired his important studies of the influence

of Indian astrology on Arabic and Persian texts and consequently on the Greek astral science tradition of Byzantium. Pingree also explored the mirror image of this transmission in an earlier era, tracing the assimilation of Hellenistic horoscopic astrology into the Sanskrit discipline of *yavana-jātaka* via the pre-Gupta Indo-Greek kingdoms in western India. This work, which became his doctoral dissertation and his first monograph, led directly to his prolonged collaboration with the great Assyriologist historian of science Otto Neugebauer, and to his study of Akkadian, Arabic and Persian in order to track similar exchanges of scientific knowledge through the entire four-thousand-year history of the great Eurasian development of the exact sciences. (Even erudition as deep as Pingree's was forced to submit to some limitations in this quest, however: to the end of his life he remained somewhat regretful, and slightly apologetic, that he had never found time to undertake learning Chinese.)

Pingree's hypotheses about cross-cultural transmission of scientific ideas were frequently groundbreaking and sometimes controversial. A few of them were tentative speculations that never attained wide acceptance, such as his suggestion that the structure of Bhāskara II's table of easy Sines (*laghu-jyā*) might reflect influence from Islamic trigonometry. Some were plausible inferences unproven by conclusive textual evidence that were accepted by most mainstream historians but strongly resisted by some others, such as his attribution of parts of the Sanskrit *Jyotiṣa-vedāṅga*/*Vedāṅga-jyotiṣa* to Babylonian mathematical astronomy techniques in the Achaemenid period or his derivation of Āryabhaṭa's planetary mean motions around 500 CE from a hypothesized Greek source. Some were preliminary conclusions based on little-known textual sources that remain to be explored more fully, such as his deductions about the relationship of Mesopotamian and ancient Indian astral omen literature or about the recognition of Islamic optics in seventeenth-century Sanskrit astronomy. But most of his discoveries concerning scientific transmission, including his study of the Indian adoption of Hellenistic astrology, his exposition of the dependence of medieval Spanish astronomy on Indian sources, and his research on the reactions to Latin heliocentrism in eighteenth-century Rajasthan, were solidly established by masterly textual scholarship and have substantially transformed the standard narrative of scientific development.

In light of the above-mentioned controversies, Pingree has sometimes been rashly relegated (by those acquainted with only a few isolated fragments of his work) to the company of reactionary Orientalists like the nineteenth-century John Bentley and G. R. Kaye who took it for granted that original scientific discoveries were by default, Greek, and that treatises in, say, Arabic or Sanskrit or Chinese must represent mere borrowings and imitations. Pingree himself, however, was far from sharing this ill-informed view, which he scornfully dismissed as the vice of "Hellenophilia" unworthy of a serious historian:

> Hellenophiles, it might be observed, are overwhelmingly Westerners, displaying the cultural myopia common in all cultures of the world but, as well, the arrogance

> that characterized the medieval Christian's recognition of his own infallibility and that has now been inherited by our modern priests of science... If it is evident that for a historian the proposition that the Greeks invented science must be rejected, it necessarily follows that they did not discover a unique scientific method... Babylonian and Indian mathematics are frequently criticized for relying not on proofs but on demonstrations. But without axioms and without proofs Indian mathematicians solved indeterminate equations of the second degree and discovered the infinite power series for trigonometrical functions centuries before European mathematicians independently reached similar results... Those who deny the validity of alternative scientific methods must somehow explain how equivalent scientific "truths" can be arrived at without Greek methods. And in their denial they clearly deprive themselves of an opportunity to understand science more deeply. (Pingree, "Hellenophilia versus the history of science", *Isis* 83.4 (1992), 554–563)

It should be noted that the "infinite power series" referred to above were familiar to Pingree largely through the Sanskrit editions of the Kerala-school treatises published by his brilliant and indefatigable colleague K. V. Sarma. The historiographic approach that they, and the studies in the present work, espouse–namely, the careful exploration of the intellectual content of a scientific tradition within its own cultural context and in its encounters with other cultures–is what David Pingree recognized as the true vocation of a historian of science.

Department of Mathematics, Union College, Schenectady NY, USA.

K. V. Sarma (1919–2005)[1]

M. S. Sriram

Born at Chengannur in Kerala on 27th December 1919, Krishna Venkateswara Sarma had his school education in Attingal near Thiruvananthapuram. He completed his B.Sc. degree with Physics as the major subject in 1940, from Maharaja's College of Science, Thiruvanathapuram. His family tradition of Sanskrit scholarship influenced Sarma to join the M.A. course in Sanskrit at Maharaja's College of Arts, Thiruvanathapuram, which he completed with distinction in 1942. During 1943–51, he was in charge of the Manuscripts Section of the Kerala University Oriental Research Institute and Manuscripts Library. It is here that he acquired expertise in deciphering and critically editing palm-leaf and paper manuscripts of Sanskrit and Malayalam texts. During this period, he prepared an analytical catalogue of nearly 50,000 manuscripts of the library.

From 1951 to 1962, Prof. Sarma was in the Department of Sanskrit, University of Madras, where he was associated with the project of compiling the *New Catalogus Catalogorum of Sanskrit Works and Authors*, under the direction of the great Sanskritist V. Raghavan. It was also the time when his life-long pre-occupation with the Kerala school of Astronomy and Mathematics began to take shape and he started painstakingly collecting manuscripts on Astronomy, Astrology and Mathematics, critically editing and translating many of them. Some of his early publications in this genre were *Grahacāranibandhana* of Haridatta, *Siddhāntadarpaṇa* of Nīlakaṇṭha, *Veṇvāroha* of Mādhava, *Goladīpikā* and *Grahanāṣṭaka* of Parameśwara. During this period, Prof. Sarma also came under the influence of the renowned scholar T. S. Kuppanna Sastri, in collaboration with whom he edited the main text of the Vākya system, *Vākyakaraṇa*, with the commentary of Sundararāja.

At the invitation of Acharya Viswa Bandhu, Prof. Sarma moved in 1962 to the Visvesvaranand Institute of Sanskrit and Indological Studies of the Panjab University at Hoshiarpur. He served as the Director of the Institute during 1975–80 and stayed on at the Institute till 1983. This period of his stay at Hoshiarpur was indeed very productive and he published more than 50 books, mostly on the

[1] This is based on the obituary which appeared in Ind. Jour. Hist. of Sci. 41(2006), 231-246, which also includes a bibliography of publications of K. V. Sarma.

Kerala School of Astronomy. These include very important seminal works such as *Dṛggaṇita* of Parameśvara, *Golasāra* of Nīlakaṇṭha, *A History of the Kerala School of Hindu Astronomy*, *Līlāvatī* of Bhāskarācārya with *Kriyākramakarī* of Śaṅkara and Nārāyaṇa, *Tantrasaṅgraha* of Nīlakaṇṭha with the commentaries *Yuktidīpikā* and *Laghuvivṛti* of Śaṅkara, *Jyotirmīmāṃsā* of Nīlakaṇṭha, and *Gaṇitayuktayaḥ*.

In 1983, Prof. Sarma returned to South India to settle down in Chennai. His important publications during this period include: *Indian Astronomy: A Source Book* jointly with the renowned historian of science B.V.Subbarayappa, *Vedāṅga Jyotiṣa* of Lagadha and *Pañcasiddhāntikā* of Varāhamihira, on which he had worked in collaboration with T. S. Kuppanna Sastri. From 1990 onwards, Prof. Sarma had been working on a critical edition and English translation of the celebrated Malayalam work *Gaṇita-yukti-bhāṣā* of Jyeṣṭhadeva (c.1530 AD). He requested K. Ramasubramanian, M.D.Srinivas and M.S.Sriram to prepare detailed explanatory notes in English. Prof. Sarma also edited a Sanskrit version of *Gaṇita-yukti-bhāṣā* which appeared in 2004, and compiled an important catalogue, *Science Texts in Sanskrit in the Manuscripts Repositories of Kerala and Tamil Nadu*, which includes a list of nearly 3,500 works related to science and technology. In fact, he continued to be relentlessly active till his very death (on January 13, 2005).

Prof. Sarma has to his credit several publications also on diverse aspects of Sanskrit learning such as Vedas, Itihāsas and Purāṇas, Dharmaśāstras, etc. In fact, he has authored more than 100 books and 500 articles. His outstanding contribution consists in searching for and bringing to light many of the seminal works of Kerala School of Astronomy, which show that the tradition of Mathematics and Astronomy continued to flourish till late middle ages at least in the South of India. They also present a detailed view of the methodology of these sciences, on issues such as justification of mathematical and astronomical results and procedures, and the importance of continuous examination and revision of planetary theories. It is mainly due to Prof. Sarma's painstaking work on primary sources that the work of the Kerala School has been brought to the attention of historians of Mathematics, and opened a new perspective on Indian contributions during the late medieval period.

Just as in the case of his illustrious predecessors such as Bibhutibhushan Datta, Avadhesh Narayan Singh and others, and his own contemporary and collaborator K.S.Shukla, Prof. Sarma did not receive even in his own country, the recognition and accolade, which he richly deserved. He was of course awarded the D.Litt degree of Panjab University in 1977 and, in 1992, was bestowed the Certificate of Honour by the President of India. He was also conferred the honorary degree of Vācaspati by Kendriya Samskrita Vidyapeetham, Tirupati, in 2003.

Prof. Sarma, with the wish that his legacy should continue, founded the Sree Sarada Education Society & Research Centre during the 90s, donating his lifetime savings and invaluable collection of books and manuscripts; since his demise,

the Centre has published critical editions of some of the texts that he had been working on during his last days. One hopes that the Centre will receive support and encouragement from scholars and funding agencies so as to sustain the torch that Prof. Sarma lit.

Department of Theoretical Physics, University of Madras.

Geometry in the Śulvasūtras

S. G. Dani

Śulvasūtras are compositions pertaining to the fire rituals performed by the Vedic Indians. The rituals involved constructions of altars and fireplaces in a variety of shapes, involving geometric theory. Some of the theory is explicitly enunciated, while some other aspects of the knowledge at that time can be inferred from the constructions. We present here an overview of the geometric ideas contained in the Śulvasūtras.

*Yajna*s, or fire rituals, formed an integral part of life in the Vedic culture, going back to 1500 BCE or earlier, and extending until about the sixth century BCE. Some of these concerned sacrifices to be performed regularly by a householder (*gṛhastha*), while performance of certain others was prescribed for bringing about fulfilment of specific aims or desires, which included both material (acquiring cows, vanquishing an enemy etc.) and transcendental (securing place in heaven) aspects.

In view of the great significance attached to the *yajna*s meticulous attention was paid to a variety of details in their planning and execution. The rituals involved construction of altars (*vedi*) and fireplaces (*agni*) in a variety of intricate shapes. Over a period the procedures for construction of the shapes appear to have got more formalised and acquired a degree of sophistication in geometrical terms. The Śulvasūtras mark the peak in the geometrisation of the altar building activity of the Vedic era.

The Śulvasūtras are often referred to as 'manuals' for construction of the altars and fireplaces. While there is a certain valid analogy here, it should be borne in mind however that the contents are not limited to prescribing steps or procedures for the construction of the altars and fireplaces. They also describe various geometric principles involved, and set up a body of geometric ideas. In Baudhāyana and Āpastamba Śulvasūtra there are separate sections devoted to geometric theory.

The Vedic people were a heterogeneous community, with many *śākhās* (branches), having nevertheless a common cultural identity. The different *śākhās* had their versions of Śulvasūtras, transmitted orally from generation to generation within the community (branch). There are nine extant Śulvasūtras of which

four, Baudhāyana, Āpastamba, Mānava and Kātyāyana Śulvasūtras are of significance from a mathematical point of view. The dates of the Śulvasūtras are uncertain, but it is generally believed that they were composed sometime during the period 800 – 200 BCE; for the individual Śulvasūtras the ranges would be, Baudhāyana (800 – 500 BCE), Āpastamba and Mānava (650 – 300 BCE) and Kātyāyana (300 BCE – 400 CE), according to Kashikar, as quoted in [11].

The root *śulv* means 'to measure', and the name Śulvasūtra would correspond to "theory of mensuration" (see [4]). The word *śulva* also means 'rope' in Sanskrit, which indeed was a major equipment employed in measurements. In the body of the Śulvasūtras the word *śulva* does not appear; instead the word *rajju* is used for rope. On the other hand, while most measurements involved in the constructions were indeed carried out with ropes, there are instances where a *bamboo* rod was used instead. This suggests that the name Śulvasūtras indeed was meant to convey mensuration in the conceptual sense, and not only as operations with the rope. The meaning of *śulva* as rope is presumably a later development.

Like other *Vedāṅgas* (appendages of the *Vedas*) the Śulvasūtras are composed in the *sūtra* (aphoristic) style, characterised by short sentences with nouns often compounded at great length and verbs avoided as much as possible, rather than running prose, presumably for reasons of convenience in reciting them. The text, which is in prose form in other respects, has been divided by later commentators into convenient segments, treated as individual *sūtras*, and grouped into Chapters. As presented in [19] (the scheme which we shall follow in the sequel for reference), Baudhāyana has 21 Chapters adding to 285 sūtras, Āpastamba has 21 Chapters adding to 202 sūtras, Mānava has 16 Chapters adding to 228 sūtras, and Kātyāyana has 6 Chapters adding to 67 sūtras.

There is a considerable overlap in the contents of the different Śulvasūtras indicating that the works are expositions from a common stream of knowledge. There are also significant differences, which may be attributed to the different branches they come from and the difference in their period. For general reference in this respect the reader is referred to [19]; (see also [1], [6], [13], [14] and [16]).

The contents of the Śulvasūtras can be broadly categorised into two groups, one consisting of Geometric theory, and another dealing with various details about the constructions of various *vedis* and *agnis*; in Baudhāyana and Āpastamba the sections dealing Geometric theory are arranged in the beginning. From the other part also one can draw some inferences about the geometric knowledge at that time. The preponderant aspect in this part however is the description of the "Architecture" of tilings involved in the construction of complex figures needed for the *vedis* and *agnis*; (Kātyāyana Śulvasūtra however consists mostly of theory part - this also explains its being shorter than the others). The designs of some of the special fireplaces involve elaborate figures resembling falcons and other birds, tortoise, chariot wheels, circular trough (with a handle), pyre, etc., whose

constructions are described, quite elaborately, in the form of tiling by bricks in certain primary shapes. Apart from the issue of achieving likeness with the desired figure, in terms of rectilinear constructions, there are also other stipulations involved, such as the number of bricks to be used etc.. On account of these there are also some arithmetical and combinatorial features involved, in a rather scattered form, in the architectural description of the layout of the fire altars. A study of this aspect would be of interest. We shall however not concern ourselves with it here, and will confine to geometry in the Śulvasūtras, including the principles and constructions described explicitly, as well as those which can be seen to be involved implicitly.

The geometric contents from the Śulvasūtras will be discussed in the following sections taking up various themes. Before going over to the main contents, a few words would be in order regarding the units of measurement involved. They had various units for measures of lengths. Measures of many of the altars are given in terms of *puruṣa* (meaning man), which was about $7\frac{1}{2}$ feet, stipulated as the height attained by the performer of the sacrifice, *yajamāna*, with uplifted arms. A commonly occurring small unit is *aṅgula* (meaning finger, in width); *puruṣa* comprised of 120 *aṅgula*s, so an *aṅgula* was about $\frac{3}{4}$ th of an inch. *pada* (meaning foot, given as 15 *aṅgula*s in Baudhāyana and 12 *aṅgulas* in Kātyāyana), *prādeśa* (12 *aṅgula*), *aratni* (24 *aṅgula*) are some of the other length measures that occur frequently.

The Baudhāyana Śulvasūtra gives in the beginning (sūtra 1.3) names of 18 different units of length measure. The smallest among them, *tila* (sesame seed) is $\frac{1}{34}$ of an *aṅgula*. It was postulated by Thibaut (see [20], page 15) that the unit owes its origin to the fact that they had a formula for $\sqrt{2}$ involving the fraction $\frac{1}{34}$; (the formula will be discussed later). Many of the intermediate units do not bear a simple fractional relation with *purusa* however; e.g. a *bāhu* is 36 *aṅgula*s, a *yuga* is 86 *aṅgula*s, etc.. The units must have arisen from the context of performance of specific *vedis* and many of them occur infrequently.

The other Śulvasūtras use many of the units described by Baudhāyana but there is no systematic listing or a comprehensive statement on their interrelations as in Baudhāyana Śulvasūtra. Some other measures, *vitasti*, *ūrvasthi*, *aṇūka* are also mentioned in Āpastamba Śulvasūtra.

For the area of rectilinear figures they had the notion as we have today. They were aware that for similar figures the ratio of the areas equals the square of the ratio of the lengths of the corresponding sides, as is clear from usage of the idea at various places; in Āpastamba Śulvasūtra there is also an elucidation of this with some examples, including with fractional sides $1\frac{1}{2}$ and $2\frac{1}{2}$ (sutras 3.6 to 3.9). The square units and the corresponding linear unit were known by the same name, the meaning being understood from the context; e.g. *puruṣa* could mean the unit of length as well as the area of the square with that length ("square *puruṣa*", so to speak), depending on whether it referred to length or area; we shall also adopt this

as a convention in the sequel when referring to the Śulvasūtras units, rather than prefixing the term "square".

1. Construction of Rectilinear Figures

Though inevitably there are some 3-dimensional features to the fireplaces described in the Śulvasūtras, the geometric ideas of significance chiefly concern planar geometry. These involve the concepts and construction of rectilinear figures such as squares, rectangles, symmetric (isosceles) trapezia and triangles, rhombuses, as well as circles, as primary figures.

The rectilinear figures sought to be drawn had a bilateral symmetry; viz. isosceles triangles, symmetric trapezia, rectangles. The east-west line served as the line of symmetry. Towards construction of these figures with prescribed sizes for the sides, the sutras principally describe steps to draw perpendiculars to the line of symmetry; these are however packaged into complete procedures for drawing the desired figures, as may be seen in some examples discussed below. The issue of drawing a perpendicular to a given line at a given point on it is addressed in the Śulvasūtras in two essentially different ways, involving the following principles (described here in modern formulation):

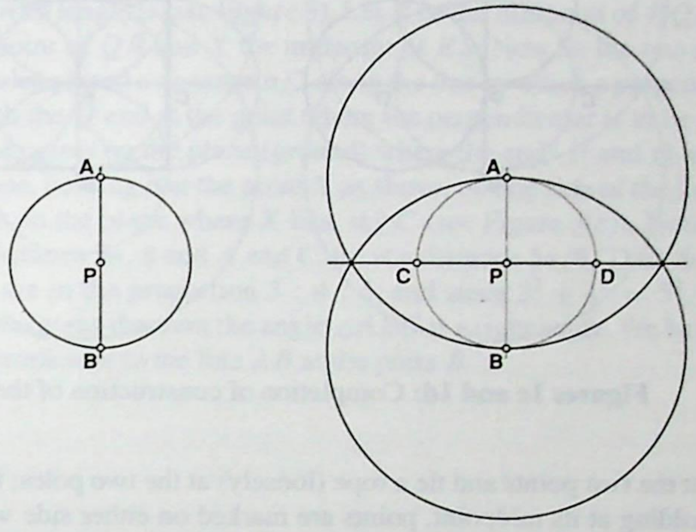
- given two circles with equal radii intersecting each other, the line joining their (two) points of intersection is perpendicular to the line joining their centres, at the midpoint of the line segment joining the centres.
- (converse of Pythagoras theorem) in a triangle with sides with lengths a, b, c if $c^2 = a^2 + b^2$ then the sides with lengths a and b are perpendicular to each other.

These principles are *not* enunciated in the Śulvasūtras, though they are implicitly at work in their constructions (see however the discussion at the end of §2). A large number of sūtras describe constructions of squares and trapezia via application of one of the above, for various specific given sizes, as we shall see in some detail below.

The first statement as above is of course what we commonly use for drawing perpendiculars, to a given line at a given point, in Euclidean geometry: take two points equidistant from the given point and draw two intersecting circles with these as centres and join the points of intersection. This procedure is involved in various constructions in Śulvasūtras for drawing perpendiculars in the same way as we now do. It is however not isolated as a procedure for drawing perpendiculars, but forms a part of the package prescribed for construction of various figures (in the framework as indicated above): thus a construction of a square in Baudhāyana (sutras I.22 to I.28 [19] consists of the following steps, described as an aggregate (see Figure 1):

i) take a rope of the desired side of the square and mark the midpoint;

ii) place a pole at the desired midpoint (say at P as in Figure 1a) and tying the ends of the rope to it draw the circle around it by the mark (at the midpoint) and place poles at the points where the circle meets the east-west line, (A and B as in Figure 1);



Figures 1a and 1b: Beginning of Baudhāyana construction of a square

iii) draw circles with centres at points A and B by the length of the rope, and mark the points where the line joining their points of intersection meets the original circle, to obtain the north-south line and mark the points where the line meets the circle as in (ii) (C and D as in Figure 1b);

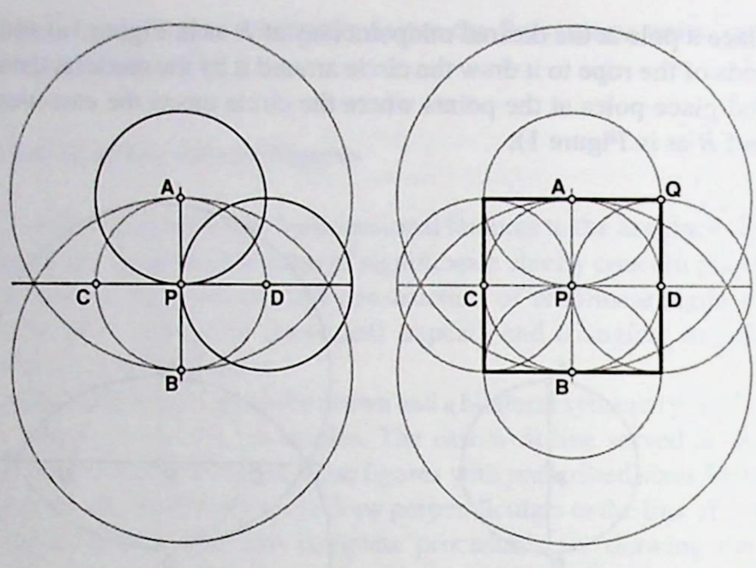
(iv) draw circles by the mid-point on the rope, tieing both its ends at A and then at D and mark the point where they meet (Q as in Figure 1d); this is one of the vertices of the square, and the others can be obtained similarly.

Notice that the first three steps are designed to draw the perpendicular bisector to the line of symmetry and marking the midpoints of the sides of the desired square, and in (iv) these are used to produce the vertices of the square.

Other constructions are also described in a similar vein, as a package for producing the intended figure, without reference to steps involved in each other.

The procedure for drawing the circles involved in the above construction is by tying one end of a rope to a pole placed at the point chosen to be the centre and tracking the point at a distance equal to the selected radius; this plays the role of the compass as used in school geometry now.

The same principle as above was also used with a rope in another way (see Figure 2): given a line and a point on it, say P, to draw the perpendicular to the line at P, take two points on the line equidistant from P, on either side of the line, affix



Figures 1c and 1d: Completion of construction of the square

poles at the two points and tie a rope (loosely) at the two poles; then stretching the rope holding at its midpoint, points are marked on either side where it lies on the ground; the line joining these two points is the desired perpendicular to the original line at P.

It may be seen that though the procedure is different the same principle underlies this construction. This variation for producing perpendiculars also appears as part of construction of some rectilinear figures; in particular Baudhāyana construction of a rectangle given in sutras I.36 to I.41 adopts this procedure.

Though these methods based on the orthogonality principle as above have been used in various constructions, on the whole the *sūtrakāra*s show greater predilection towards using the other principle, namely the converse of Pythagoras theorem. It is believed that the Egyptians also used triangles with sides 3, 4 and 5 (in some

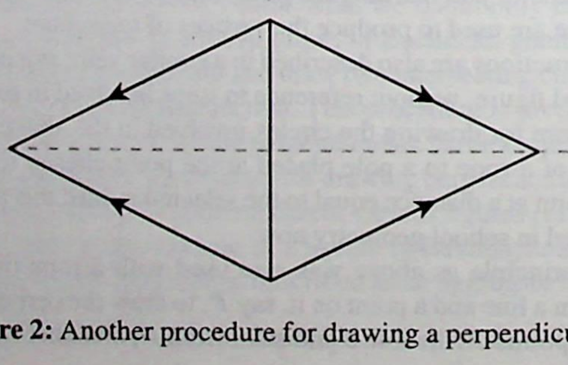
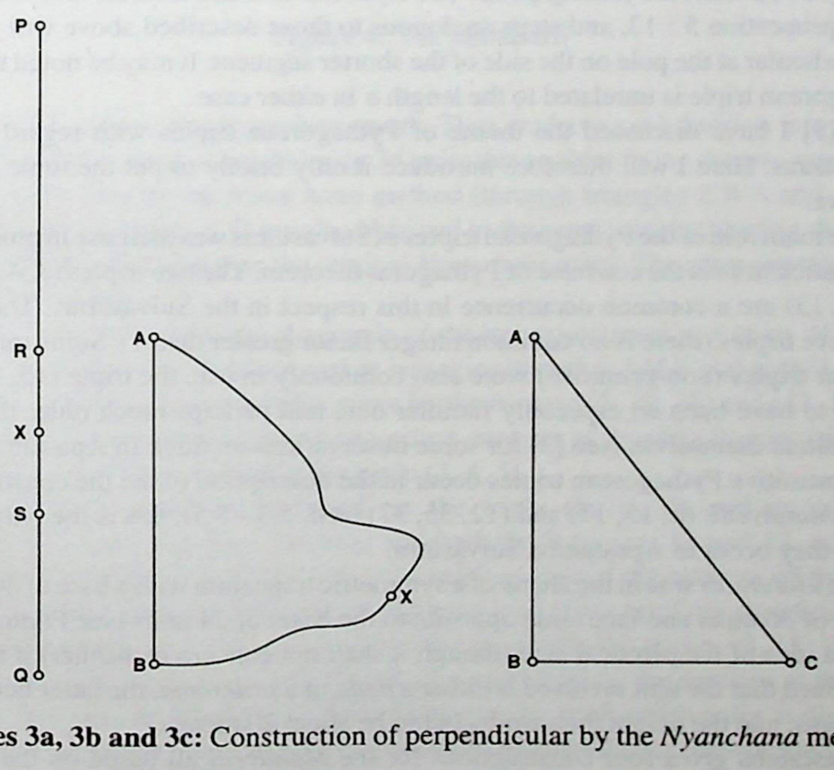


Figure 2: Another procedure for drawing a perpendicular

units) for construction of perpendiculars, but this has now been discounted (see [8], Appendix 5). It is not clear whether there is any other instance historically of using converse of Pythagoras theorem for the purpose of drawing perpendiculars.

Let me now describe in some detail how this application was made, and discuss the possible convenience for which it was preferred. Take a rope, with endpoints marked P and Q, with length a (see Figure 3). Let R be the midpoint of PQ (on the string), S the midpoint of QR and X the midpoint of RS. Now tie the two ends of the string to two poles placed a distance $a/2$ along the line to which a perpendicular is to be drawn, with the Q end at the point where the perpendicular is to be drawn. Let A and B be the points on the plane (ground) where the ends P and Q are tied. Now stretch the rope, holding it at the point X as above, on one side of the line AB, and mark the point on the plane where X lies, say C (see Figure 3(c)). Notice that B and C are at a distance $3a/8$ and A and C are at a distance $5a/8$. Thus the sides BC, AB and AC are in the proportion $3 : 4 : 5$, and since $3^2 + 4^2 = 5^2$, by the converse of the Pythagoras theorem the angle $\angle ABC$ is a right angle. We have thus constructed a perpendicular to the line AB at the point B.



Figures 3a, 3b and 3c: Construction of perpendicular by the *Nyanchana* method

A rope with markings as above, which can be preserved, can thus be used as an *instrument* to draw a perpendicular, essentially at one stroke; the distance $a/2$ to be kept between A and B is also available by a mark on the rope as the distance

between P and R. In directing a *yajamāna* towards drawing a perpendicular (as a step in the construction of the *vedi*), it would be simpler for the priest to use this approach, than those with the orthogonality principle discussed above.

The above procedure depends on the fact that a triangle with sides in the proportion 3 : 4 : 5 is a right angled triangle, by the converse of the Pythagoras theorem. The Śulvasūtras describe also analogous procedures using in place of (3, 4, 5) other triples (a, b, c) such that $a^2 + b^2 = c^2$; typically they are triples of integers, that we now call *Pythagorean triples*, and their multiples by a fraction, but occasionally some incommensurable triples are also involved. As in the above procedure it involves marking a point X so that when the rope is stretched holding at that point we would get a right angled triangle. Such a point is called *Nyanchana*; *Nyanchana* means "lying with face downwards" and in this context signifies that the marked point is to be plotted on the ground. A procedure involving use of the triple (5, 12, 13) goes as follows: having chosen a distance a between the poles, a rope of length one and half times the measure is taken (thus extending the rope by $a/2$), and the *Nyanchana* mark is set at a distance a sixth of the extended piece, namely $a/12$ from the joining point. The *Nyanchana* mark then divides the string in the proportion 5 : 13, and steps analogous to those described above will yield a perpendicular at the pole on the side of the shorter segment. It may be noted that the Pythagorean triple is unrelated to the length a in either case.

In [3] I have discussed the theme of Pythagorean triples with regard to the Śulvasūtras. Here I will therefore introduce it only briefly to put the topic in perspective.

The main role of the Pythagorean triples in Śulvasūtras was their use in producing perpendiculars via the converse of Pythagoras theorem. The two triples (3, 4, 5) and (5, 12, 13) are a common occurrence in this respect in the Śulvasūtras. These are primitive triples (there is no common integer factor greater than 1). Some multiples of these triples (non-primitive) were also commonly in use; the triple (15, 36, 39) seems to have been an especially familiar one, and perhaps much older than the Śulvasūtras themselves (see [3] for some observations on this). In Āpastamba two more primitive Pythagorean triples occur in the description of the the construction of the *Mahāvedi*: (8, 15, 17) and (12, 35, 37) (Asl. 5.3 - 5.5); this is the only place where they occur in Āpastamba Śulvasūtra.

The *Mahāvedi* was in the shape of a symmetric trapezium with a base of 30 units, height of 36 units and face (side opposite to the base) of 24 units (see Figure 4); to give an idea of the physical size (though it shall not concern us further) it may be mentioned that the unit involved is either a *pada* or a *prakrama*, the latter being $\frac{1}{4}$th of *puruṣa*, and the height then works out to be about 20 meters.

Āpastamba gives four constructions for the *Mahāvedi* all based on the *Nyanchana* method as discussed above. The first one known as the *ekarajjuvidhi* ("one-rope process") involves a rope of length 54 units, with markings at 36 and 12 units from the two ends respectively and a mark at the midpoint of the remaining middle

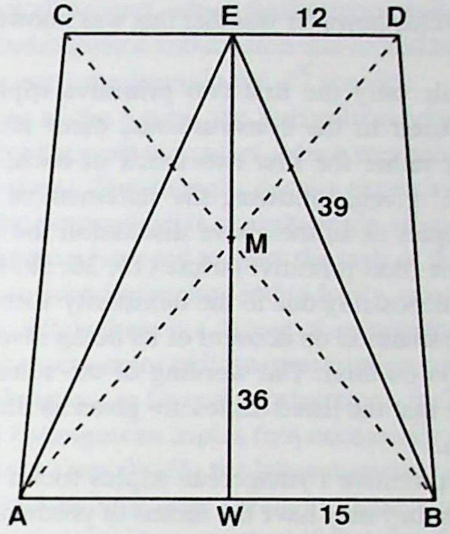


Figure 4: The *Mahāvedi*

portion of 6 units as the *Nyanchana* mark. Thus we have a subdivision into 15 and 39 units, and the desired height being 36 units this enables fixing the two vertices A and B at the base by the *Nyanchana* method (through triangles EWA and EWB. The vertices at the face CD are also obtained in the same way, by plotting the mark at 12 from the end rather than the original *Nyanchana* mark. The other constructions involve two cords.

It may be noticed that the diagonals of the trapezium meet at a point M which together with the base and face makes isosceles triangles AMB and CMD whose symmetric half parts have lengths given by the triples (15, 20, 25) and (12, 16, 20) respectively. Āpastamba's second procedure consists of drawing the two triangles using that these are multiples of the triple (3, 4, 5).

The third construction is based on the triple (5, 12, 13) and its multiple (15, 36, 39), using that half the base and half the face, viz. 15 and 12, occur in these triples; unlike in the first two constructions one of the vertices of the right angled triangle involved (the one on the east-west line) is of no significance to the diagram itself. It is the last construction in which the triples (8, 15, 17) and (12, 35, 37) appear. Again since 15 and 12 occur in these, the vertices at the base and the face may be plotted with these triples, respectively, with poles at distances of 8 and 35 from the base and face respectively. Manifestly the construction would no longer be as elegant as the earlier ones, and Āpastamba could not have missed noticing that. From the overall context it is clear that the aim has been to quench the curiosity about the various ways the *Nyanchana* method could be used to plot the vertices, and no practical requirement is involved. This involves finding triples

with 15 and 12 as one of the first two entries. Incidentally, Āpastamba exhausts such triples. There is no clue however whether this was known, and if so how it was realised.

In Baudhāyana, while only the first two primitive triples, viz. $(3,4,5)$ and $(5,12,13)$, are found used in the constructions, there is a list of 5 primitive Pythagorean triples (or rather the first two terms of each, which of course determines the third term) given, following the statement of Pythagoras theorem. Apart from the four triples as in the above discussion the triple $(7,24,25)$ also forms part of the list; the (non-primitive) triple $(15,36,39)$ is also included along with the others, which is possibly due to the familiarity with it in a wider context, and its association with tradition on account of its being involved in the Mahāvedi (see [3] for a discussion on this). The wording of the sūtra, its location and the overall context indicate that the listed triples are given as illustrative examples for the Pythagoras theorem.[1]

There are no other primitive Pythagorean triples found in the Śulvasūtras. It would seem that though they may have the means of producing more triples, if not an infinite family, they would have had no motivation for it, with the ones that occur having specific objectives that are adequately met in their context.

The *Nyanchana* procedure was also used in a construction with an incommensurable triple (not a multiple of Pythagorean triples); the construction is also interesting from another point of view (see below). To construct a square of side a Āpastamba gives the following procedure: on a rope mark three points, two endpoints and a point in between whose distance from one endpoint is $a/2$, and from the other endpoint it is equal to the length of the diagonal of the square with side $a/2$. The rope is now tied to two poles, one at the desired centre of the square and the other at the point at distance $a/2$ on the east-west line, by the endpoints as above, with the longer side (from the in between mark) being attached to the centre. The rope is then stretched, holding it at the middle marked point. Where it lays on the ground is one of the vertices of the desired square; the other vertices are plotted similarly. The procedure presupposes being able to mark a point at a distance equal to the diagonal of a square of side $a/2$, so the construction is in a way "circular", from the point of view of logical development. However, the diagonal of a square was such a common occurrence in their practice that it was I suppose treated as a "tangible" quantity. Having once constructed a square one could produce ropes with markings as required, and then employ them for later construction of squares following the above procedure. It has been suggested by some authors (see Footnote 4) that an approximate numerical expression for $\sqrt{2}$ was used to get the

[1]Indeed, a Pythagorean triple does not, strictly speaking, illustrate Pythagoras theorem, since one would need to know that the triangle corresponding to the triple as side-lengths is a right angled triangle; the nuance of the illustration here would be more like "when you draw rectangles with sides 3 and 4, 5 and 12, . . . , the diagonals will be 5, 13, . . . , and you see that the area produced by the diagonal is the sum of the that of the squares on the sides."

diagonal of the square with side $a/2$; this however seems unlikely, considering the common usage of ropes all around, which furthermore would give a more accurate measure, without the cumbersome subdivision that would be involved in producing the approximate value using the formula for $\sqrt{2}$ (see §6).

For the constructions of the figures the Śulvasūtras adopted meticulous procedures, involving drawing the perpendiculars, which was accomplished by the methods that we discussed above. Surprisingly, despite extensive use of these methods, conceptualisation of the perpendicular or right angle seems to have eluded them. Absence of the concept may have led to view the task of drawing each symmetric trapezium with different given dimensions individually, as is noticeable especially in Āpastamba Śulvasūtra. While there is a degree of unity in the descriptions, with the concept of the perpendicular a more uniform prescription could have been given for the constructions. Furthermore in several constructions, including in the *Mahāvedi* as seen above, special Pythagorean triples (not necessarily primitive) were sought depending on the desired sizes. While the latter may have offered an amusing diversion, in practice it would have been simpler to have a unified way of drawing perpendiculars, even with *Nyanchana* method if that was found more convenient, and marking the point on the perpendicular line at the desired distance, producing the line if necessary. This possibility does seem to have been realised at some stage. In [3] I have noted that the *vedis* described in the Asl. 6.3-6.4 (*nirūḍhapaśubandha vedi*), Asl. 6.6, Asl. 6.7 (*paitṛki vedi*), Asl. 6.8 (*uttara vedi*), and Asl. 7.1 involve varied shapes (two trapezia of different dimensions, two squares and an oblong rectangle), but the construction of each of them refers to "Having stretched (the cord) by the mark at fifteen" (*pancadaśikenaivāpāyamya*), and taking various contextual factors into account concluded that the phrase is used as a way of saying "Having drawn a perpendicular"; the desired point on the perpendicular line is meant to be marked on that line by measuring out the requisite distance. This marks a step towards conceptualisation of the perpendicular at a practical level, which however does not seem to have been abstracted further.

2. Pythagoras Theorem and its Applications

The most notable feature of the Śulvasūtras in terms of geometric theory is the statement of the so called Pythagoras theorem. This stands out especially in the context of the fact that some of them, especially Baudhāyana, predate Pythagoras. There has been a variety of speculation in this respect, including that Pythagoras may have got it from the Indians (A. Bürk quoted in [10]) or, in broader terms, that there may have been a common source for the geometry of the Greeks and the Indians (see Seidenberg [17] and [18]). Available inputs seem inadequate to have a meaningful discussion on this, and in any case we will not go into this aspect here. The main discussion below will pertain to the role of the theorem in the overall context of the Śulvasūtras themselves.

Śulvasūtras do not contain proofs, and there are no indications of how the statements were arrived at; in this respect the analogy with manuals is pertinent. The idea of proof in the sense of Euclidean geometry does not really apply to the context of the Śulvasūtras, since they were essentially concerned only with the practical aspects, and at best one can ask how they would have reached the individual conclusions, say in terms of practical geometry. While in the case of many statements the knowledge may be attributed to geometric intuition, it is difficult to see how they may have come to the statement of something like the Pythagoras theorem. We will discuss some of the possibilities that have been proposed in this respect. At the end of the section I propose a new hypothesis on how the Pythagoras theorem may have been thought about in the context of the Śulvasūtras.

Neither the notion of a right angle nor of a right angled triangle are found in the Śulvasūtras, as concepts; of course right angled triangles appeared as parts of various figures, and were implicit in the *Nyanchana* operations, but were not identified separately. Thus statement of the Pythagoras theorem occurs not with respect to right angled triangles, but rather with reference to rectangles. A close translation of how it is stated in Baudhāyana would be "the diagonal of a rectangle makes as much (area) as the (areas) made separately by the base and the side put together". An identical statement is found in Āpastamba. The statement appears also in Kātyāyana where it is followed by a clause "*iti kṣetrajnānam*", meaning "this is the knowledge about areas". In Mānava however an equivalent statement occurs in quite a different form. A close translation would be "multiply the stretch by the stretch and the width by the width; the square root of the sum is how much the diagonal is". The difference is striking. While the former presents the statement about *areas*, the latter describes it as an algorithm for computing the *length* of the diagonal. Together with the fact that many geometric constructions described in the other Śulvasūtras are missing from Mānava, this suggests a shift from a geometric approach to either algorithmic or functional approach. We shall however see later that Mānava also contains some interesting original geometric statements not found in the other Śulvasūtras.

The Pythagoras theorem is also applied in the Śulvasūtras towards the following *theoretical* objectives.

To describe a square whose area is the sum of areas of two given squares: This is covered in two parts, one when the given squares are of equal size, another when they are unequal. For the former the answer is given as the diagonal of the square. For the latter the prescription is to mark the side of the smaller square on one of the sides of the larger square (from one of the vertices) and take the diagonal of the rectangle formed by it with the other side of the bigger square (Figure 5; PC is the prescription for the side of the sum of the squares with sides PA and PB). The square on this diagonal has area equal to the sum of the two given squares. This is of course a direct application of the Pythagoras theorem.

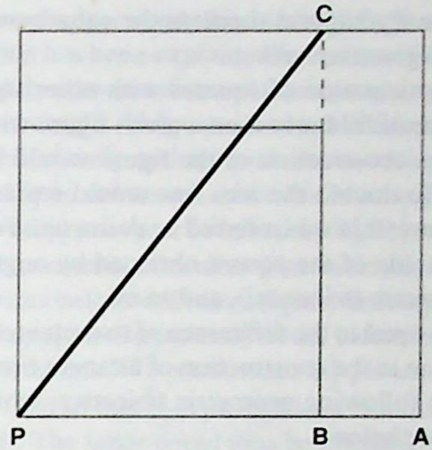


Figure 5: Square with area equal to the sum of two given squares.

To describe a square whose area is the difference of areas of two given (unequal) squares: For this again one marks the side of the smaller square in the larger one from a corner, forming a rectangle as before. An arc is drawn with one of the vertices of the rectangle as the centre and the longer side as the radius, meeting the opposite side of the rectangle. The portion of that side that is cut out, from the vertex nearer the point picked as the centre, is the answer (see Figure 6 ; BC as in the figure is the prescription for the side of the difference of the squares with sides PA and PB).

It may be seen that the cut out portion together with the smaller side of the rectangle are sides of a rectangle whose diagonal equals the side of the bigger

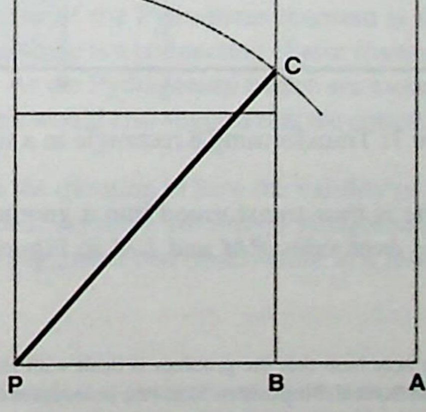


Figure 6: Square with area equal to the difference of two given squares.

square, and hence by the Pythagoras theorem the square over it has area equal to the difference of the two given squares.

The summing or augmentation of squares with other squares was involved in producing figures with multifold area than a given figure with the same shape. For this the procedure in the construction of the figure would be followed as before, with an enlarged unit. To double the area one would replace the unit by the size diagonal of the unit square; this was referred as *dvikaraṇi*. For tripling the area one would replace it by the side of the square obtained by augmenting a square with area two with the unit square (*trikaraṇi*), and so on.

The square with area equal to the difference of two squares does not seem to have had any direct application in the construction of altars or fireplaces. It was however involved in meeting the following geometric objective, which in turn was needed in practice as we shall see below.

To transform a rectangle into a square with the same area[2]: For this purpose, forming the square over the smaller side, the remaining rectangle is divided into two equal parts parallel to the square and the farther part is then taken off and put adjacent to the other side of the square (see Figure 7).

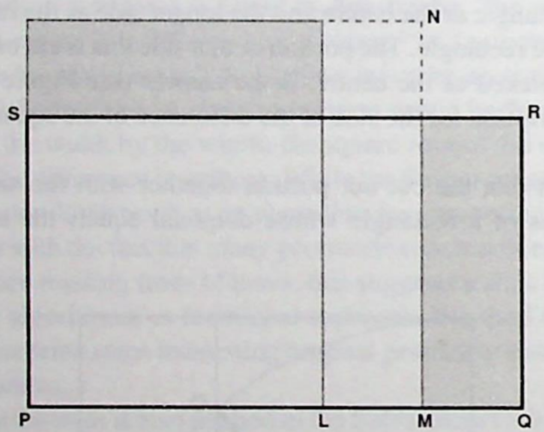


Figure 7: Transforming a rectangle to a square

The original rectangle is thus transformed into a gnomon, viz. a difference of two squares (the squares over sides PM and LM in Figure 7). How to transform

[2]It may be worthwhile to note here that the problem is dealt with Euclid's Elements in Book II, Proposition 14, based in turn on Book II, Proposition 5, and the procedures described by Baudhāyana and Euclid closely correspond to each other, though the presentation differs in certain details. It would appear that they would have drawn the figures somewhat differently, but this is a matter of speculation.

this into a square is already described; this is how they put it also (!): *tasya nirhāra uktah*, "resolution of this has been explained", so states Baudhāyana.

As rectangles could thus be transformed into squares one could in turn augment a square with a given rectangle. This was involved in the following. Various *yajnas* were to be repeated periodically and for each successive performance, the area of the fireplace (*agni*) had to be enhanced, while maintaining its shape. In many cases the first fire altar had area $7\frac{1}{2}$ *puruṣas*, followed by $8\frac{1}{2}$, $9\frac{1}{2}$, and so on, and it could go upto $101\frac{1}{2}$ *puruṣas*[3]. To construct a fire altar with an enhanced area, one would carry out the same procedure as before with the original unit replaced by an appropriate larger one. This would apply independently of the (often intricate) shape of the fire altar. To increase the area from $7\frac{1}{2}$ *puruṣas* to $8\frac{1}{2}$ *puruṣas*, for instance, the unit would have to be increased so that $\frac{2}{15}$th *puruṣa* area would be added to each *puruṣa* in forming the new unit. The latter could thus be obtained as the side of the square combining a square *puruṣa* with a rectangle of area $\frac{2}{15}$th *puruṣa*.

It may be recalled that the converse of the Pythagoras theorem is also a simple consequence the theorem itself. Use of the converse in the constructions as seen in the last section may also be viewed as an application of the Pythagoras theorem. Actually it would seem that the *sūtrakāras* viewed the converse as a part and parcel of the theorem; clubbing the converse with the direct statement is not uncommon in the course of intuitive thinking. Otherwise it is difficult to explain why they would make it a point to state the direct theorem (as is done in all the Śulvasūtras) but not the converse that is much more used. There are some other points which seem to corroborate this. Following the statement of the Pythagoras theorem in Āpastamba Śulvasūtra there is a clause *tābhirjneyābhiruktam viharaṇam*. This translates as "the above mentioned constructions are knowable (deducible) from this". The constructions referred to are of squares by the *Nyanchana* method, using the triples (5, 12, 13) and (3, 4, 5) respectively, which involves the *converse* of the Pythagoras theorem. Thus the converse seems to be viewed as an integral part of the theorem itself, though not stated as such. Similarly, as mentioned earlier, in Baudhāyana the statement of the Pythagoras theorem is followed by a list of the Pythagorean triples, and there is a connecting clause *ityetāsūpalabdhiḥ*, translating as "this is obtained in". As the Pythagorean triples are more closely connected with the converse theorem this would also suggest that the converse was clubbed together with the theorem itself.

Let me now come to the question of how the validity of the Pythagoras theorem may have been concluded? As there are some Pythagorean triples occurring in the Śulvasūtras it has been suggested that observation in a few cases was extrapolated

[3] This is according to [19] and [11]. It however seems to this author that the original sutra translates as going upto $108\frac{1}{2}$ *puruṣas*, adding one *puruṣa* 101 times; this of course has no consequence to any mathematical discussion.

into a general statement (see [21]). This explanation overlooks the point that knowing a Pythagorean triple is not tantamount to knowing a right angled triangle (or rectangle); a Pythagorean triple would be of significance only *after* one has arrived at the Pythagoras theorem, and not lead to it. Datta [4] (see also [19]) has proposed possible geometrical arguments they may have known. The connection of the construction involved however seems rather distant from the Śulvasūtras constructions for the argument to be convincing.

The following possibility does not seem to have been considered in literature. From the context of the Śulvasūtras we see that they faced the question as to how the squares add up (rather than the other way question as to what is the size of the diagonal, or the square over it)? Augmentation of squares, for the purpose of the working out the size of the enhanced unit as explained earlier, was in all likelihood the motive behind the discovery of the theorem, as has been noted already by Seidenberg [18]. If one were to contemplate on the question in the old times prior to the theorem being discovered, with the conviction of a mystic that there has to be a natural choice, it seems natural that the diagonal of the rectangle with the two sides of the two squares (perhaps the one emanating from the common point when the squares are set on the two sides of the rectangle) would pose as a natural candidate to give the same area[4]; what else can it be? Experience with square tilings would also aid such an intuition. Having guessed the answer it could be confirmed in various ways; it does not even have to be only with triangles with rational sides, though the latter would indeed facilitate verification in some ways.

3. Transforming a Square into a Circle

Towards the end of the purely geometric part Baudhāyana gives (in sutra I.58) a geometric construction to produce a circle with same area as that of a given square. Essentially the same procedure is described in the other three Śulvasūtras, though the wording in each of them varies a little, including in usage of the geometric terms.

Finding a circle with a given area, for which the above procedure would be used, was involved in the construction of fire altars in the shape of a chariot wheel (*rathacakraciti*), a circular trough (with a handle) (*droṇaciti*), rounded tortoise (*parimaṇḍala kūrmaciti*).

Given a square, the prescription for obtaining the desired circle goes as follows. Take half the diagonal of the square and drop it from the centre along the midriff, and draw the circle including a third of the part jutting out (see Figure 8; QR is $\frac{1}{3}$ rd of QS, and PR is the prescription for the radius of the desired circle). For a square with side $2a$ the prescribed radius works out to be $a + \frac{1}{3}\left(\sqrt{2} - 1\right)a = \left(2 + \sqrt{2}\right)a/3$.

---

[4]This may be compared with realising that the resultant of two forces at a point represented by two vectors is given by the vector represented by the diagonal of the parallelogram formed by the two.

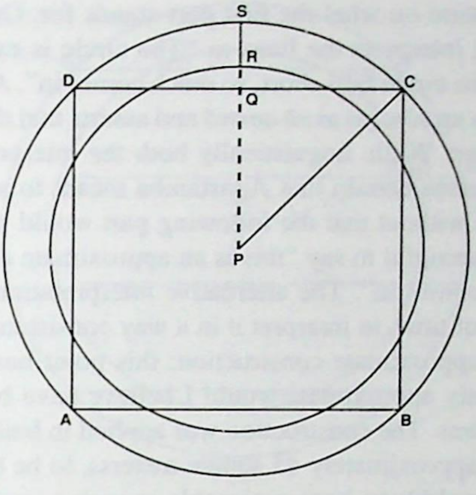


Figure 8: Circling the square.

The area of this circle for the unit square works out to be 1.0172524 . . . , slightly more than that of the original square.[5]

In Baudhāyana there is no further comment on the sūtra. On the other hand a sutra in Āpastamba (see the next paragraph) says in particular "as much is left out that much comes in", the reference being to the portions of the square that are left out when the circle is drawn and those which are incorporated inside. The comment gives an insight into how the choice for the radius of the circle was arrived at. Observing that the diameter should be between the length of the side and that of the diagonal, they looked for a proportion of the extra part which would ensure that as much area is left out that much comes in. The correct proportion for this (in the light of the modern knowledge) turns out to be 0.30993473 . . . , and if it is to be approached by a simple fraction, as one of certain number of parts, then it would have to be one in three. That is the number they picked, presumably from intuitive considerations. Incidentally, if they had chosen the proportion to be 3 out of 10 (not inconceivable in the context of the number 10 having acquired significance as a base in counting numbers) they would have got a better result; for the unit square they would have got a circle with area 0.99271948 . . . .

The sutra from Āpastamba relating to the area of the circle is followed by *sānityāmaṇḍalam* and then *yāvaddhī yate tāvadāgantu*. The second part means "as much is left out that much comes in", which was quoted above. There has been

[5]In the literature there is an inclination to talk of something like "what value of π this amounts to", if the area of the circle is to be 1 as expected (it turns out to be 3.0883118 . . .). I find this peculiar; there was no concept of π in the Śulvasūtras, and they were not computing such a ratio. The comparisons should be in terms of the output produced, as followed above.

a debate in the literature on what the first part stands for. One of the commentators Karavindasvāmi interprets the lines as "The circle is exactly as large as the square, for as much the circle falls short, so much comes in". Another commentator Kapardisvāmi breaks up *sānityā* as *sā anityā* and asserts that the circle construction is an approximate one. While linguistically both the interpretations are possible for the first part it seems certain that Āpastamba meant to state that the circle is exactly as large, and without that the following part would be lacking in context. It does not seem meaningful to say "this is an approximate construction; as much is left out so much comes in". The alternative interpretation is motivated by the call, after a passage of time, to interpret it in a way consistent with the knowledge that it is in fact an approximate construction; this point has been made in [20]. The fact that it is only approximate would I believe have been suspected in the times of the Śulvasūtras. The construction was applied in building fire altars of the size of $7\frac{1}{2}$ *puruṣa*s (approximately 45 square meters), to be tiled into 200 tiles. A close to 2 % error would have been noticeable in such a context to some of those directly involved with the altar building. On the other hand, the error is not big enough to make a definitive case about it. As I see it, Āpastamba's statement is aimed at putting forth forcefully, in the face of doubts (murmurs ?) of sceptics, the scholastic position of the time: that is the exact circle; as much is left out so much comes in. In course of time the suspicions would have gained weight, serving as motivation to look for an alternative construction, and one was indeed found by the time of Mānava Śulvasūtra, which we discuss below. In Kātyāyana however only the original construction is described; even though it is a later Śulvasūtra, being in a different stream it was perhaps untouched by the development around that sutra.

Mānava describes another construction for the circle with the area of the given square (cf. stra 11.15 in [19] and 10.3.2.15 in [11]). While it actually happens to be more accurate as we shall see below, the statement is very unclear, as a result of which it does not seem to have been understood properly; in particular I believe the comment on it in [19][6] is unwarranted, and the algebraic presentation in [11] is incorrect. Let me first present my interpretation of what the sūtra means, and then discuss the wording in the sūtra and the lacunae making it unclear, that apparently led to the confusion over it.

Given a square, draw a line parallel to one of the sides and dividing the square into one-third and two-third parts (see Figure 9). Extend the line to meet the circle through the vertices of the square. Now consider one of the segments of the line from the square to the circle and take the point at one-fifth of its length, from the point on the square. Draw the circle passing through this (with centre at the midpoint of the square); in Figure 9 QR is $\frac{1}{5}$ th of QS and PR is the prescription

[6]The comment is: "Possibly these are not problems of quadrature of the circle. Ordinary squares are drawn without any mathematical significance."

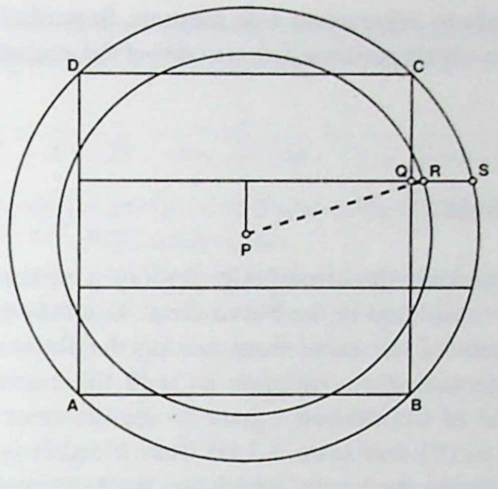


Figure 9: Mānava construction for circling the square.

for the radius of the desired circle. For a square of side length 2 the length of the segment between the square and the circle is seen to be $\left(\frac{\sqrt{17}}{3} - 1\right)$, so the square of the radius of the circle is $\left\{1 + \frac{1}{5}\left(\frac{\sqrt{17}}{3} - 1\right)\right\}^2 + \frac{1}{9}$, and this yields the area of the circle to be 3.9787159 . . . ; for the unit square the area of the prescribed circle will be 0.99467897 . . . , a much more accurate value compared to the earlier one.

The wording of the sūtra however has a lot to be desired from the point of view of clarity. Firstly, rather than asking the square to be cut into three equal parallel strips the sūtra asks it to be divided into 9 equal squares, via trisection along both the sides; presumably this is on account of a preoccupation with symmetry. Secondly, there is no mention of the outer circle. This seems to have led the authors of [19] to give up on the meaning of what follows, and content themselves presenting only an ad hoc translation, with the comment as noted. In [11] the author has been prescient of the unmentioned outer circle and that a portion of the segment between the square and the circle is involved. But here there is confusion in interpreting the second half of the sūtra. This part has a clause *lumpetpurīṣeṇeh* which translates as "cover it with loose earth", meaning obliterate it (a part of the segment as above), but there is ambiguity about what part is to be obliterated; it could refer to obliterating $\frac{1}{5}$ th part keeping the rest, or the other way around. In [11] the former interpretation is adopted, which may be appropriate linguistically, but is not what is actually intended. From the whole context it is clear that the remaining $\frac{4}{5}$ th part of the segment was meant to be obliterated; it would be too much of a coincidence otherwise for the value to come out as close as seen above. It may be borne in mind that at the time of the Śulvasūtras these shortcomings in the presentation would not have mattered, as the correct meaning would have been handed down orally. The

text was intended only to serve as an aide memoir. In particular the unmentioned outer circle would have been presumed on account of the analogous one in the earlier construction.

4. Squaring the Circle

The problem of "squaring the circle" viz. finding a square with the area of a given circle is also considered in the Śulvasūtras. Baudhāyana Śulvasūtra, which has the best treatment of the issue from among the Śulvasūtras, deals with it, not in terms of a geometric construction as with the converse, but by relating numerically the side of the desired square to the diameter of the given circle; it has been argued in [9] that sutra 3.2.10 from Mānava is in fact a geometric construction for squaring the circle, which we shall discuss later in the section. Baudhāyana gives two formulae in this respect, the second being qualified as an approximate one. The first of them we shall discuss in some detail below. The second consists of taking $\frac{13}{15}$th of the diameter of the given circle for the side of the desired square. Curiously, only this more approximate formula has been described in Āpastamba and Kātyāyana Śulvasūtras for the purpose. This formula is very crude: for the circle of unit radius it gives the square with side $\frac{26}{15}$, whose area would be $\left(\frac{26}{15}\right)^2 = 3.0044444\ldots$, in place of the correct value $3.1415927\ldots$, smaller by more than 4%. Even in terms of fractions with small denominators there are better approximations possible. The Egyptians took $\frac{8}{9}$th of the diameter as the side of the corresponding square (see [8], Chapter 13). This gives for the unit circle the square with area $3.1604938\ldots$, which is more than the actual value by just about 0.6%. Another simple fraction, $\frac{7}{8}$, as the ratio of the side of the square to the diameter also yields a better result: for the unit circle the area of the square would be 3.0625, which is about $2\frac{1}{2}$% less than the actual value. Incidentally, $\frac{7}{8}$ would have been a natural choice for Baudhāyana, in giving an approximation, since as we shall see below it is a component of the refined value given by Baudhāyana. It would seem that $\frac{13}{15}$ was the traditional value (going farther back in time) and even after the refined formula was discovered the traditional value continued to be quoted without modification.

In the construction of the altars there does not seem to be any occasion to convert a circle into a square. This would suggest that the relation as above was recorded in the Śulvasūtras (all four) only to convey a general sense of magnitude (and perhaps in deference to tradition). On the other hand the other value described in Baudhāyana, that was referred to above and we will now discuss, is apparently determined with serious effort.

The refined formula for the ratio of the side of the square to the diameter of the circle with the same area is described as follows: divide the diameter into 8 parts,

divide one of them into 29 parts and remove 28 of the parts and from the remaining one remove a sixth part and include eighth of it; thus it is

$$\frac{7}{8}+\frac{1}{8\times 29}-\frac{1}{8\times 29\times 6}+\frac{1}{8\times 29\times 6\times 8}.$$

The fraction is $\frac{9785}{11136}$ and for the circle with unit radius yields the area 3.0883265 . . . , a little more than 98.3% of the actual value.

A discerning reader may notice that Baudhāyana's construction of the circle with the area of a given square produced a circle whose area was bigger by about 1.7% while the above prescription in the opposite direction produces a square which is smaller in size by about 1.7%. In other words if one started with a square followed the construction as in the last section to get a circle and then used the above formula to get back the square it would be very close to the original. A more precise computation (in place of what is pointed out as something to be noticed at first glance) shows that if one starts with the unit square, transforms it into a circle and then transforms that into a square as indicated, one gets a square with area 1.0000024 . . . (!). This is no coincidence. The desired fraction as above was evidently computed as the inverse of the ratio involved the other way, namely in circling the square, or else we could not have got such a close answer. If the computation of the inverse could have been exact the above number would have come out to be exactly 1. However in the computation of the inverse certain approximations had to be made, leading to the small difference.

It may be recalled that for the unit square the diameter of the corresponding circle obtained by the Baudhāyana method would be $\left(2+\sqrt{2}\right)/3$. For $\sqrt{2}$ Baudhāyana has a(an approximate) formula, which we shall discuss in detail in the next section, associating with it the value $1+\frac{1}{3}+\frac{1}{3\times 4}-\frac{1}{3\times 4\times 34}$, and the ratio $\left(2+\sqrt{2}\right)/3$ to be inverted would thus be $1+\frac{1}{9}+\frac{1}{9\times 4}-\frac{1}{9\times 4\times 34}$ (expressed in the style of the Śulvasūtras). By our current method we would write the number as $\frac{1393}{1224}$ and produce $\frac{1224}{1393}$ as the inverse. However, this method may not have been known at that time. Thus they either did not identify the number as $\frac{1393}{1224}$ or did not consider the reciprocal $\frac{1224}{1393}$ for the inverse formula, for one or other reason, including possibly that it can not be expressed in terms of making smaller number of parts as in the above expressions, as 1393 is a prime! Instead their computation seems to have gone along the following lines. A part $\frac{7}{8}$ was first separated as the significant part; this could be either following a way that they may have had to find the inverse of a fraction expressed as a sum of the terms as above, or could be an ad hoc choice noticing that the inverse should be greater than that. The remaining part after separating out $\frac{7}{8}$ turns out, in our notation, to be $\frac{41}{1393\times 8}$. They however opted to replace it by $\frac{41}{1392\times 8}$, which could be split up as $\frac{1}{8\times 29}-\frac{1}{8\times 29\times 6}+\frac{1}{8\times 29\times 6\times 8}$, thus arriving at the quadrature formula.

It does not seem to have been noted in the earlier literature, but it turns out that taking only the first two terms in Baudhāyana expression yields a better result. The fraction would then be $\frac{51}{58}$, which for the circle with unit radius yields the area 3.0927467 ..., closer to the actual value $\pi = 3.1415927\ldots$ compared to the value 3.0883265 ... obtained above. Thus the additional two terms in fact worsen the result in terms of the correct value for the area of the circle. Of course this could not have been anticipated by Baudhāyana. Interestingly, however, if the alteration of 1393 in the denominator to 1392 were done before separating out the part $\frac{7}{8}$ then they would have 'accidentally' got the inverse to be $\frac{51}{58}$ as above. This however is not as close to the inverse of $\left(2+\sqrt{2}\right)/3$ as Baudhāyana's value was seen to be, which would have been the objective at that time.

The choice of $\frac{7}{8}$ as the part to be separated is interesting, especially if it was ad hoc. Why was it not chosen instead to be $\frac{13}{15}$, the supposed approximate value. Was it known that it was better than the latter, even though it had not been stated? Why did it then continue to be unstated even after the refined formula was exposited. Or were they from two different substreams of thought, even though they occur alongside each other in Baudhāyana?

Let me now come to the sutra from Mānava Śulvasūtra, sūtra 3.2.10, to which a reference was made in the beginning of the section. Some developments around it are rather puzzling. Though the sutra is rather condensed, an overall reading readily indicates that it has something to say on the question of squaring the circle, as also the converse. However various commentators, including van Gelder who produced the first edited version and English translation, have interpreted it to relate to only the converse problem of circling the square, in a way equivalent to that in Baudhāyana. A new interpretation was given by Hayashi in [9], after some emendation of the original text which has been justified on grounds of rectification of possible corruption of the manuscript over a period. According to [9] both the problems of circling the square and squaring the circle are addressed in the sutra. The square with the area of a given circle is described to be of side equal to the height of the equilateral triangle over the diameter of the circle; the converse statement continues to be interpreted to be essentially as in Baudhāyana, but there is also a slight variation offered (attributed to the referee of the paper). It may be observed that for the circle with unit radius the side-length of the stipulated square would be $\sqrt{3}$, which yields the area of the square to be 3, in place of $\pi = 3.1415927\ldots$, less by about 4.5%. This is a rather crude value, much cruder than even the approximate value given in Baudhāyana and stated also in Āpastamba and Kātyāyana. In the context of the improved result from Mānava Śulvasūtra that we saw in the last section, both direct and converse statements as above seem surprising; the first part on account of the crudity of the value and the second being not as sharp as the improved result that we noted. However such anomalies seem to be inherent in Śulvasūtras, as we have seen elsewhere also, since they continued to exposit the

traditional values of their time with no critical evaluation, even after better results became known. It may be noted in this context that the *sūtrakāra* characterises the contents of the above sutra, by a mention in the preceding sutra, as "prescriptions of informed persons", traditional knowledge in other words. The refined method for circling the square described in the preceding section comes later (though it is not clear if much can be read from the sequencing) and is presumably either due to the author himself or more contemporary knowledge of his time.

While we are discussing the area of circle the reader may wonder what the Śulvasūtras had to say about the circumference of the circle. In Baudhāyana there is only a incidental reference to it, where a circular pit "with diameter 1 *pada* and circumference 3 *padas*" is mentioned, indicating that the circumference was taken to be 3 times the diameter; interestingly this coincides with the value according to the Bible. In Mānava Śulvasūtra however the circumference is stated to be $3 + \frac{1}{5}$th of the diameter (sutra 10.3.2.13 as in [11]; the interpretation of the sutra in [11] as describing the side of the square is incorrect); the statement of the ratio is followed by "not a hair-breadth remains"! Evidently a lacuna in the classical value had been overcome and it was worth exulting over it!

5. The Square Root of 2

Three of the four Śulvasūtras, Baudhāyana, Āpastamba and Kātyāyana, describe the following formula for $\sqrt{2}$ (in words) expressing the value as

$$1 + \frac{1}{3} + \frac{1}{3 \times 4} - \frac{1}{3 \times 4 \times 34}.$$

The sūtra is followed by *saviśeṣaḥ* in Baudhāyana, *sa viśeṣaḥ* in Āpastamba and *sa viśeṣa iti viśeṣaḥ* in Kātyāyana. The word *viśeṣaḥ* means "extra" and the wording in Kātyāyana Śulvasūtra conveys that the number is in excess of the desired quantity (cf. [12]). *viśeṣaḥ* was also used as a technical term for the excess of the diagonal over the original measure, and *saviśeṣaḥ* thus signified "together with the extra", namely the diagonal itself. As such the term itself is not connected with the value being considered to be in excess, or approximate.

Though we would not know how they obtained it, such a complex expression would have been arrived at by some process (and not directly by intuition or guesswork). Whatever the process, along the way it would have been noticed that the targeted magnitude has not been reached exactly. Thus it would have been known, at least at the time when the value first came to be assigned, that it was not exact. (It is also tempting to argue that they could have squared the number and found that it is not 2, but it is doubtful if they would have taken up such an arithmetic exercise).

In decimal expansion the value of the expression is 1.4142157 . . . , which is remarkably close to the actual value 1.4142136 . . . . The closeness is striking. A few points should be noted in this respect however. Firstly, it may be recalled that the Babylonian cuneiform tablet YBC 7289 (Yale Babylonian collection), dating back to around 1800 to 1600 BCE also contains an approximate value for $\sqrt{2}$, in the sexagesimal system as 1; 24; 51; 10, which yields the $1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3} = 1.4142129\ldots$. Furthermore the Babylonian value described in a pre-existing standardised system with uniform subdivisions, namely the sexagesimal system of representation, carries an intrinsic sense of the level of accuracy. The Śulvasūtras value however is evidently the output of a specific process, and there is no intrinsic gauge of accuracy involved. No details of how the procedure went are found in the Śulvasūtras but some likely methods have been proposed, that we shall discuss below. In the context of these it would seem that while refining on the approximation $1 + \frac{1}{3} + \frac{1}{3\times4} = \frac{17}{12} = 1.4166667\ldots$, the next step took them directly to the above fine value, due to certain special features of the number $\sqrt{2}$; it could thus have been fortuitous to have got so close to the actual value.

Actually there is also no indication that they had serious interest in accuracy of that level. No other nonexact relation found in the Śulvasūtras is that close. It may be recalled that even when better values were available in the case of the quadrature formula, for instance, only a rough approximation was recorded in the later Śulvasūtras. For $\sqrt{2}$, Mānava, long after Baudhāyana, proffers only the approximate value $\frac{17}{12}$. Thus while they were deeply preoccupied with geometric aesthetics, high numerical accuracy is hardly seen to be considered seriously as an issue to be pursued.

There have been various suggestions on how the formula may have been arrived at. One convincing possibility put forth by B. Datta [4] (see also [19]) goes as follows. The presentation here is essentially along the lines of [19] with some modifications for clarity. We have two unit squares to be put together and formed into a square. Divide the second square into three equal parts by lines parallel to one of the sides. Two of these you attach at the sides of the first square sideways; this produces a gnomon (see Figure 10). Divide the third part into a square of side $\frac{1}{3}$ and a rectangle of sides $\frac{2}{3}$ and $\frac{1}{3}$. The square part is used to convert the gnomon into a square of side $1 + \frac{1}{3}$. The rectangle is divided into four equal parts along the length, into rectangles with sides $\frac{2}{3}$ and $\frac{1}{3\times4}$. Lengthwise two of these strips can be placed along the sides of the square formed earlier, side by side. We thus get a square except for a missing little square at the corner, of side $\frac{1}{12}$.

Now what we need is a square which is a difference of a square with side $1 + \frac{1}{3} + \frac{1}{3\times4} = \frac{17}{12}$ and a small one of size $\frac{1}{12}$. If strips of width $\frac{1}{12\times34}$ are cut out along the two sides, that would adjust for the square to be removed, except for a tiny square with side $\frac{1}{12\times34}$ formed of the overlap of the two strips. The latter is ignored, and the square formed by cutting out the strips is produced as the answer.

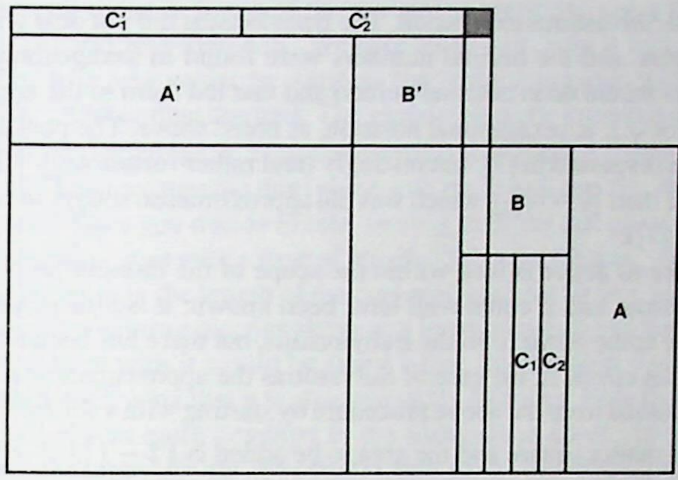


Figure 10: Possible route to the formula for $\sqrt{2}$.

It has been argued in [7], convincingly, that the Babylonian value of $\sqrt{2}$ (noted above) would have been obtained by a procedure involving the following idea. Suppose we have a square with a given sidelength and we would like to get a square with an area larger by a certain amount which is relatively small compared to the area of the original square. The difference between the desired new square and the original one is seen to consist of two (long) rectangles and a small square (the reader may see Figure 6 for reference); the long side of the rectangles has the original sidelength and the other side corresponds to the desired increment. As the remaining square is small, the desired increment in the area divided by twice the original sidelength would be a good approximation for the requisite increment in the length; the area to be added is thus put in the two rectangles themselves. In modern notation, if l is the original sidelength and α is the amount of area to be added, then $l + \frac{a}{2l}$ is an approximation to $\sqrt{l^2 + \alpha}$; this may be compared with the binomial expansion. Similarly, if we have a square of a given sidelength and wish to get a square whose area is smaller by a certain amount, then the latter divided by twice the length of the original square gives is an approximation to the requisite reduction in length; this corresponds to $l - \frac{a}{2l}$ as an approximation to $\sqrt{l^2 - \alpha}$.

It is illustrated in [7] in detail how the Babylonians may have arrived at their value of $\sqrt{2}$ using the idea, in two steps: Starting with a square with sidelength $\frac{3}{2}$ and seeking to get a square with area 2, the side is reduced by $\left(\frac{9}{4} - 2\right) / \left(2 \times \frac{3}{2}\right) = \frac{1}{12}$, to $1\frac{5}{12}$ (in [7] the details are explaiend in the Babylonian sexagesimal notation, but we shall not go into it here). This is still a larger value and a second reduction is effected to get a second approximation to $\sqrt{2}$. The fraction that one is after for reduction in the second step would be (in our notation) $\left(\left(\frac{17}{12}\right)^2 - 2\right) \times \frac{6}{17} = \frac{1}{3\times4\times34}$, same as the

last term in the Śulvasūtras expression. The Babylonians did not deal with fractions directly however, and the desired numbers were found in sexagesimal points (as our calculators would do in decimal points) and that led them to the approximation 1; 24; 51; 10 for $\sqrt{2}$, in sexagesimal notation, as noted above. The possible routes to this have been discussed in [7]. Interestingly (and rather fortuitously!) 1; 24; 51; 10 is closer to $\sqrt{2}$ than $\frac{17}{12} - \frac{1}{12\times 34}$ which was the approximation sought to be computed (according to [7]).

A procedure as above is well within the scope of the thought process involved in the Śulvasūtras, and it could well have been known; it is also possible that the idea may have come down from the Babylonians, but there has been no supporting evidence to that effect. In the case of Śulvasūtras the approximation as above may have been obtained from the above procedure by starting with a square of sidelength $1\frac{1}{3}$. This is a smaller square and the area to be added is $\left(2 - \left(\frac{4}{3}\right)^2\right) \times \frac{3}{2\times 4} = \frac{1}{3\times 4}$, which yields the first approximation $\frac{17}{12}$. In the second step this would lead to the quoted value, as seen above. This seems to fit better with the way the number is expressed.

It has been pointed out in [7] that the approximation as above is *mathematically* equivalent (emphasis as in original) to Heron's method for extraction of square roots, which involves the following: Let $A = l^2 + \alpha$. If l is a (first) approximation to $\sqrt{A}$, A/l also approximates $\sqrt{A}$, and moreover if $l < \sqrt{A}$ then $A/l > \sqrt{A}$ and vice versa. Therefore $\frac{1}{2}\left(l + \frac{A}{l}\right)$ is candidate for a better approximation. But $\frac{1}{2}\left(l + \frac{A}{l}\right) = l + \frac{1}{2l}(A - l^2) = l + \frac{\alpha}{2l}$, the approximation seen earlier. It may therefore seem tempting that to argue that the approximate expression for $\sqrt{2}$ was obtained via such averages. However, the arriving at it in this way involves a kind of familiarity with algebra that is not evidenced in the Śulvasūtras; [7] does not seem to suggest it in respect of the Babylonians either.

Some authors have also brought in the more elaborate expression $l + \frac{\alpha}{2l} - \left(\frac{\alpha}{2l}\right)^2 / 2\left(l + \frac{\alpha}{2l}\right)$ for $\sqrt{l^2 + \alpha}$ ([15] or [19], and [2]) as explanation for the approximations for $\sqrt{2}$ in Śulvasūtras. Some of the other suggestions, including one by Thibaut, who produced the first English edition of Baudhāyana supplementing it with his commentary, also involve taking difference of squares (see [20] or [19]), relying on some arithmetic. These explanations however do not seem natural for the context, and seem to depend on ideas that evolved only in a later period, unlike the geometrical explanations above which are consistent with Śulvasūtras style.

Here is another possibility which I would like to propose, taking the context of the Śulvasūtras into account. Having found the need for a numerical value for the diagonal of the square (I will argue below that this was the case), one may simply choose to find it by measurement with the ropes. This seems natural also from the point of view that they were using ropes for a variety of constructions. Take a rope of length equal to the diagonal of a square. Mark out one unit from one end. By

measuring the unit piece in terms of the remaining piece one notes that the latter is more than $\frac{1}{3}$rd and less than half. We take away $\frac{1}{3}$rd from it and consider the remaining part. It is now easier to compare the remaining piece with the second piece of length $\frac{1}{3}$, rather than the unit. The remaining piece turns out to be slightly less than $\frac{1}{4}$th of the second piece. Thus the diagonal is slightly less than $1 + \frac{1}{3} + \frac{1}{3 \times 4}$. How much less? You measure the third piece with the length that fell short. It is very close to 34 times. Thus you decide to take away $\frac{1}{34}$th of the last piece, which gives the formula. One may start with a rope of length 35 feet for the unit, so the required total rope length, namely the length of the diagonal, is about 50 feet. In this case the length of the final remaining piece as above will be over an inch and the penultimate piece to be measured with it would be about 3 feet, comfortable sizes for manual measurements, establishing that it is almost exactly 34 times. Note that use of over 50 feet long ropes was quite common in the Śulvasūtras constructions; the rope involved in the *ekarajjuvidhi* for the construction of the *Mahāvedi* measured over 100 feet. There are some variations possible over the above theme, and though one may ask why this one was picked, the answer can simply be that through trials this would have been found most convenient.

What was the motivation of the *sūtrakāra*s for determining the value of $\sqrt{2}$? We have seen before that the $\sqrt{2}$ is involved in the Śulvasūtras as the diagonal of the unit square. It is called *dvikaraṇi*. As a magnitude it was involved in doubling the area of the figures to be drawn and, as mentioned earlier, this was achieved by enhancing the reference unit to the size of the diagonal of the corresponding unit square, and following the same steps of construction as before, for the figure to be drawn. For this purpose however one does not actually need the numerical value of $\sqrt{2}$. Nor would the complex formula be of much help in producing a mark for *dvikaraṇi*, as it would involve the cumbersome process of subdividing the segments into as many as 34 parts!.[7] It would be much simpler to draw a square and mark its diagonal on the rope. The numerical value was needed on the other hand in deriving the quadrature formula, in order to find the inverse of $\left(2 + \sqrt{2}\right)/3$, as seen in the last section. As noted by Seidenberg [17] almost certainly this problem served as the motivation for finding a numerical value for $\sqrt{2}$. Of course, as the reader may recall from the earlier sections, the later Śulvasūtras, Āpastamba and Kātyāyana included the value of $\sqrt{2}$ but not the refined quadrature formula as in Baudhāyana for which

[7] In [10] it is stated (on page 361), referring to the papers of Albert Bürk on Āpastamba Śulvasūtra, that the approximate formula for $\sqrt{2}$ was used for the construction of a square. The square that is referred to is in fact the one described towards the end of Section 1, constructed by the *Nyanchana* method with the triple $(1, 1, \sqrt{2})$. For the length $\sqrt{2}$ the original sūtra asks to use the *saviśeṣaḥ*. This apparently has been interpreted by these authors as using the *approximate value* for it than the diagonal itself. However *saviśeṣaḥ* stood for diagonal as well, and there is no reason to suppose that in the above construction they followed the cumbersome procedure involving the formula for $\sqrt{2}$ to fix the length, rather than using the diagonal itself, in the manner I indicated.

the value of $\sqrt{2}$ was used. This could however be simply due to the fact that once the value was found it may have seemed appealing to the later authors.

There has been unwarranted speculation in the literature, imputing knowledge of the irrationality of $\sqrt{2}$ to the *sūtrakāras* (see [4], the discussion on p. 195). They would have been aware that the value they gave was an approximate one, and that is the closest it gets. To infer from this the knowledge of irrationality of $\sqrt{2}$ is to miss or to trivialise the significance of the notion of irrationality. Irrationality is not just about observing that attempts to express a number as a fraction yield only approximate results (even this can not be said to be subsumed in its entirety in the Śulvasūtras as there is but one approximate value described), but to be able to assert, with valid argument, that it is *impossible* to express it as a fraction. The Śulvasūtras were concerned with practical aspects of geometry of altar building, and some of it they abstracted, developing a theoretical understanding. One does not find them getting involved with philosophical or arithmetical aspects of numbers, and there is no point in indulging in wishful thinking about it.

Acknowledgement

The author would like to thank Bill Casselman for producing soft copies of the figures appearing in the article. The author would also like to thank the referees for their comments on a preliminary version of the paper, and one of the referees especially for pointing out the paper of Fowler and Robson [7], on which some of the discussion in § 5 is based.

References

[1] A.K. Bag, Mathematics in Ancient and Medieval India, Chaukhambha Orientalia, Varanasi and Delhi, 1979.

[2] R. Cooke, The History of Mathematics. A Brief Course. Second edition. Wiley-Interscience (John Wiley and Sons), Hoboken, NJ, 2005.

[3] S.G. Dani, On the Pythagorean triples in the Śulvasūtras, Current Sci. 85 (2003), 219–224; (available at: http://www.ias.ac.in/currsci/jul252003/contents.htm/)

[4] Bibhutibhusan Datta, Ancient Hindu Geometry: The Science of the Śulba, Calcutta Univ. Press, 1932; reprint: Cosmo Publications, New Delhi, 1993.

[5] Bibhutibhusan Datta and Avadhesh Narayan Singh, History of Hindu Mathematics; A Source Book, Parts I and II, Asia Publishing House, 1962; reprint: Motilal Banarasidas, Delhi, 2001.

[6] Bibhutibhusan Datta and Avadhesh Narayan Singh, *revised by* K.S. Shukla, Hindu Geometry, Indian J. of History of Math. 15 (1978), No. 2, 121–188.

[7] David Fowler and Eleanor Robson, Square root approximations in old Babylonian mathematics: YBC 7289 in context, Historia Math. 25 (1998), 366–378.

[8] R.J. Gillings, Mathematics in the Time of Pharaohs, Dover Publications, New York, 1972.

[9] Takao Hayashi, A new Indian rule for the squaring of a circle: Mānava Śulvasūtra3.2.9–10, Ganita Bharati 12 (1990), 75–82.

[10] Thomas Heath, Translation of "The Thirteen Books of Euclid's Elements" with Commentary, Second Edition Vol. I, Dover Publications, 1956 (Originally published by Cambridge University Press, 1926).

[11] Raghunath P. Kulkarni, *Char Shulbasūtra* (in Hindi), Maharshi Sandipani Rashtriya Vedavidya Pratishthana, Ujjain, 2000.

[12] Kim Plofker, Mathematics in India: 500 BCE - 1800 CE, Princeton University Press, 2008.

[13] Satya Prakash and Ram Swarup Sharma, Baudhāyana Śulbasūtram, with Sanskrit Commentary by Dvarakanatha Yajvan and English Translation and Critical Notes by G. Thibaut, The Research Institute of Ancient Scientific Studies, New Delhi, 1968.

[14] Satya Prakash and Ram Swarup Sharma, Āpastamba Śulvasūtra, with commentaries from Kapardswamin, Karvinda and Sundararaja and English translation by Satya Prakash, The Research Institute of Ancient Scientific Studies, New Delhi, 1968.

[15] M.L. Rodet, Sur les méthodes d'approximation chez les anciens, Bull. Soc. Math. France, 7 (1879), 159–167.

[16] T.A. Saraswati Amma, Geometry in Ancient and Medieval India, Motilal Banarsidas, Delhi, 1979.

[17] A. Seidenberg, The ritual origin of geometry, Archive for History of Exact Sciences (Springer, Berlin) Vol. 1, No. 5, January 1975. (available at:
http://www.springerlink.com/content/r6304ku830258l85/)

[18] A. Seidenberg, The origin of mathematics, Archive for History of Exact Sciences (Springer, Berlin) Vol. 18, No. 4, December 1978. (available at:
http://www.springerlink.com/content/p3k6201x14q7685m/)

[19] S.N. Sen and A.K. Bag, The Śulbasūtras, Indian National Science Academy, New Delhi 1983.

[20] G. Thibaut, The Śulvasūtras, The Journal, Asiatic Society of Bengal, Part I, 1875, Printed by C.B. Lewis, Baptist Mission Press, Calcutta, 1875.

[21] B.L. van der Waerden, Geometry and Algebra in Ancient Civilizations, Springer Verlag, 1983.

S.G. Dani
School of Mathematics
Tata Institute of Fundamental Research
1 Homi Bhabha Road, Colaba
Mumbai 400 005
India
E-mail: `dani@math.tifr.res.in`

On the Origins of Zero

Frits Staal

1. Indic Origins?

IT IS WIDELY believed that zero originated in Indic Civilization but the evidence in support of that belief is not only meager; it is almost zero. No place or time, let alone the name of a discoverer or inventor, has ever been suggested. How can we handle such a problem? We must start from the beginning.

Indic Civilization starts with the Indus Civilization which is earlier than the Vedas. Its inscriptions exhibit occurrences and sequences of circles that resemble the numerals that have expressed zeroes in more recent times; to be a little more precise: more than three thousand years later. Other civilizations roughly contemporary with the urban complexes of Mohenjo-daro and Harappa used circles also, but they did not refer to zeroes. On cuneiform tablets from Uruk in Mesopotamia, dating to 3,000 BCE, circles refer to the number "2" (Tropfke 1980: 29). It weakens the suggestion that the Indus circles were expressions of zero. A more serious difficulty is that the language of the Indus inscriptions is not known. It is not even clear that it was a language and its uses are controversial. Some of these topics are discussed in Staal 2008: 7-11, which provides the evidence and further references.

With Vedic mathematics we are on firmer ground. We find not only geometry but integers, a rudimentary decimal system for counting. It did not include negative numbers, but "addition, subtraction and perhaps multiplication of whole numbers" (Hayashi 2003: 360-61). The Rigveda made use of recursion (Divakaran, *forthcoming*). It did, moreover, distinguish between cardinal and ordinal numbers (Renou: 1964: 92; Staal 2007: 589-590; 2008: 272-273). In all these cases we are dealing with numbers, not with numerals: the Vedas are an Oral Tradition since there was no writing on the subcontinent prior to the Buddhist Emperor Aśoka who reigned from 268 to 231 BCE. But "zero" did not only lack a symbol. There was no term for it in the oral tradition. The word *kha*, which Indian mathematicians used later to denote zero, occurs in Vedic only in the senses of "hole", "opening", "vacancy" or "space".

Counting boards based upon the decimal system took another step but the Indic evidence is of later date and the empty spaces are zeroes of a kind, not symbolic expressions. Even today, the Indian *paṇḍita* or traditional scholar uses neither

counting boards nor books: he carries his knowledge in his head. Writing had its commercial uses, manuscripts go back to the beginnings of the Common Era, but it began to be used widely only after the invention of Hinduism in the early nineteenth century (Staal 2008a).

Fortunately, it is not the end of our story since those who look for the origins of something, even if it is zero, must look beyond the domain where it is customarily located, even if it is absent. The remaining parts of my essay attempt to do so. Part 2 will discuss two pioneering investigations. Parts 3 and 4 will take us beyond the history of mathematics. Part 3 will pay attention to linguistics and Part 4 to the Vedic theory of ritual, not included in modern classifications or curricula though regarded during the period of middle Vedic as an exact science. Having gone that far we must recall, that "exact" and "science" are often no more than labels and that all names of disciplines are due to us, not to the universe to which we belong.

2. Khmer and the Buddhist Madhyamaka

Two original contributions, undigested legacies from the twentieth century, have to be taken into account if we wish to understand how zero may or may not have been discovered. The first is due to Joseph Needham, the famous scholar and scientist who published, together with his collaborators, the many volumes of *Science and Civilization in China.* The second is due to David Ruegg, a brilliant Sanskrit scholar whose early work dealt with the Sanskrit grammarians but whose chief contributions since have been to Buddhist Studies, primarily as expressed in Sanskrit and Tibetan sources.

Needham presented his ideas in the third volume of his series. It is entitled "Mathematics and the Sciences of the Heavens and the Earth" and was first published in 1959. It is important to understand what Needham tried to do for he has been criticized and misunderstood like other pioneers. He did not examine Chinese sciences from a simple "evolutionist" concept of history, as if they were "more or less clumsy attempts to express modern scientific ideas," a notion that Pingree wisely rejects in another context (Pingree 2003: 45). But neither did Needham present Chinese sciences from "the Chinese point of view," whatever that is, as Seyyed Hossein Nasr (1968: 21) tried to do with respect to the Islamic Sciences. Needham's perspective is different and he has expressed it in unambiguous terms: "to write the history of science we have to take modern science as our yardstick—that is the only thing we can do—but modern science will change and the end is not yet" (Needham 1976: xxxi with further discussion in Staal 2006: 91-97).

D. J. de Solla Price, a historian of science at MIT, described Needham's work as follows: "In my estimation, the essential contribution made thus far by the six volumes of *Science and Civilization in China* lies in the systematization and presentation in English translation or summary of the substantive content of the

otherwise ill-digested bulk of Chinese scientific and technical literature. Here we have the raw material on which generations of later scholarship can be founded. Here at last we have some map to tell us where to look, and some indication at least of what we shall find" (1971: 17-18).

Needham's volumes deal with much more than the Chinese sciences. They abound in references to Indian, Mesopotamian and European disciplines. The discussion on the origin of zero begins to meander in Volume 3 on page 9. Needham knows, of course, that zero is widely believed to have originated in Indic Civilization, he is familiar not only with the Chinese but also with Indic counting boards and then takes an unexpected turn: he zeroes in on early South East Asian inscriptions. The thesis that emerges will not surprise us but is not formulated in a few simple sentences or in a single paragraph. We get the idea when we combine three separate sentences that occur on pages 10-12 and that I quote here because they are Needham's own words:

- "The usual view is that the circular symbol for zero derived directly from India, where it first appears on the Bhojadeva inscriptions at Gwalior dated +870."
- "While the first epigraphic evidence for the zero in India is, as has just been mentioned, of the late +9th century, it has been discovered about two centuries earlier in Indo-China and other parts of south-east Asia. This fact may be of much significance."
- "It would seem, indeed, that the finding of the first appearance of the zero in dated inscriptions on the borderline of the Indian and Chinese culture-areas can hardly be a coincidence."

I shall not discuss the first sentence though more recent discoveries of Indian inscriptions have pushed the dates further back. It neither affects Needham's thesis, nor the origins of zero as we shall see. The second sentence is based upon an early article published in 1931 by another celebrated scholar, George Coedès, the French historian of South East Asia and especially of the Khmer Empire of what are now Kampuchea and parts of Thailand. He did not only write the classic account of the 'Indianization' of Indochina and Indonesia that was first published in French, often reprinted since 1964 and translated into English as *The Indianized States of Southeast Asia*. He also published eight volumes of Cambodian inscriptions, and wrote much else that inspired thousands of more recent publications, all of them supplementing, improving and updating his work, just as de la Solla Price had predicted of Needham's series. As for the third of Needham's three sentences, its significance depends on the significance of the second.

Needham's idea is based upon two inscriptions discussed in Coedès article. One comes from Cambodia, the other from Indonesia. Both show the zero, one in the form of a dot and the other in the shape of a small closed curve that may be

a circle. Both are dated 683 CE, almost two centuries earlier than the alleged first Indic inscription of 870.

Needham does not comment on the puzzling circle, if it is a circle, which comes from the small Indonesian island of Banka, but he stumbles on both inscriptions and then makes his fatal lapse. He assumes throughout that their language is Southeast Asian but does not say which language it is. Is he thinking of Khmer or Old Javanese? The simple truth is that the language of the inscriptions is expressed in a Southeast Asian script but is none other than Sanskrit.

This was obvious to Coedès as every careful reader of his article can see. He writes on page 325: "au Cambodge, les premières inscriptions *sanskrites* datées font usage des mots symboliques"; "au Champa, les deux plus anciennes inscriptions *sanskrites* datées . . .en langue *sanskrite*"; "les inscriptions *sanskrites* du Champa"; "à Java, la plus ancienne inscription *sanskrite* datée fait usage des mots symboliques" (all italics mine). These quotes show that "the first appearance of the zero in dated inscriptions on the borderline of the Indian and Chinese culture-areas" does not only fail to be a coincidence. They demonstrate that the content of these inscriptions is Indic.

I now come to David Ruegg who wrote a brief but substantial article partly concerned with our problem. It is entitled "Mathematical and Linguistic Models in Indian Thought: The Case of Zero and *Śunyatā*" and was published in 1978. *Śunyatā* refers, as is well known, to the Buddhist concept of 'emptiness.' It is a characteristic feature of the Madhyamaka school and was foreshadowed by a certain Bhadanta Vasumitra who may have lived at the end of the first or beginning of the second century CE. Its context is the theory of *dharmas*, which do not refer to the Buddhist *dharma* or 'doctrine,' but to 'elements or factors, each of which is considered to bear its own specific characteristic that determines it' as Ruegg explains the expression of the *Abhidharmakośa: 'svalakṣaṇadhāraṇād dharmaḥ'*.

I shall not further some readers' possible annoyance with Vasumitra's Sanskrit but here is Ruegg's translation: "A *dharma* evolving in the [three] times is stated to be other according to the different states it enters, [the change in question being then] due to otherness of state (*avasthāntaratah*) but not of substance". The words and phrases within square brackets are due to Ruegg, who explains the example that Vasumitra adds as follows: "like a marker or counter (*vartikā*) in reckoning which in the unit position has the value of a unit, in the hundred's position that of a hundred, and in the thousand's position that of a thousand"—a straightforward expression of the use of zero as a place-marker in the decimal system.

Ruegg adds that the same idea is sometimes expressed by the term *gulikā*, "ball" or "bead" which, like the counting boards to which they belong, should remind us of the fact that mathematicians are not always concerned with what modern readers think of almost exclusively, viz. *writing*. Needham is familiar with this usage of *śunya* because he compares it to the empty spaces on Chinese counterboards in a long footnote on his pages 11-12.

Ruegg discusses two terms for zero: *kha*, which we have already met, and *bindu*, which means "dot". He draws attention to the *Vāsavadattā*, a literary work of the sixth century by Subandhu which uses *śunyabindu* to denote the symbol for zero. He refers next to Pingree's work on the *Yavanajātaka* of Sphujidhvaja, then about to be published. That text was composed in 149/150 CE and used the term *bindu* in "the earliest reference known to the decimal place-value system with a symbol for zero in India" (Pingree 1978, II: 406, I: 494).

The next topic Ruegg considers is the history of the term *śunya.* He starts with the Rigveda which employs *śuna* for "lack, absence, emptiness." (One of the earliest examples is "lack of sons" in RV 7.1.11.) Later Vedic has *śunya* in the meaning "hollow, deserted". After providing more information on *śunyatā*, he concludes cautiously that we cannot trace connections between the Buddhist "emptiness" and the mathematical concept of zero.

Ruegg then turns to early linguistics and Pāṇini's lopa. We shall look at its apparent invisibility in the next section. Ruegg notes its occurrence in modern linguistics but then diverges from our topic in grammatical and philosophical directions that involve *śunyatā* without throwing light on the zero. His cautious peregrinations have inspired my own meanderings. The reader should note what has not been shown and remember Pingree's statement: the earliest written reference to the decimal place-value system with a symbol for zero in India is dated to 149/150 CE. We have to look not only beyond writing but further.

3. Zeroes in Sanskrit Grammar

Almost all Indic mathematicians wrote in Sanskrit, the classical language of science that unites the subcontinent (Staal 1995). Malayalam is among the famous exceptions (Divakaran 2007). Persian and English became more common in premodern and modern times but Sanskrit continued in mathematics and Jyotiḥśāstra or astronomy-cum-astrology (see, e.g., Minkowski 2002, 2008). Many of the classics of Indian mathematics were composed in concise and sometimes elegant Sanskrit verse. Here is young Āryabhaṭa on the subdivisions of time:

"A solar year is a year of men. Thirty of them make an ancestral year. Multiplied by twelve is explained as a year for the gods"(*ravivarṣaṃ mānuṣyaṃ tad api triṃśadguṇaṃ bhavati pitṛyam| pitṛyaṃ dvādaśaguṇitaṃ divyaṃ varṣaṃ samuddiṣṭaṃ*: 3:7) Another couple of lines condenses the full sine table in one couplet (1:12). And who does not know the penultimate verse: "From the ocean of true and false knowledge I have, through the boat of my own knowledge, rescued with the grace of the deity the precious sunken jewel of true knowledge" (*sadasajjñānasamudrāt samuddhṛtaṃ devatāprasādena | sajjñānottamaratnaṃ mayā nimagnaṃ svamatināvā*)

How did these mathematicians know Sanskrit? It could not have been their first or native language. Many were not brahmans (Āryabhaṭa was not or else

his name would have been Āryabhaṭṭa[1]); and neither need all brahmans know Sanskrit. *Ārya*, moreover, does not refer to "the three twice-born classes" as it probably did in the *Manusmṛti*, "The Law Code of Manu", 2.207 (Olivelle 2004: 242). The distinction between *ārya* and *anārya* has not been a racial or ethnic distinction. Madhav Deshpande has shown that it expresses claims to moral, social and spiritual status, tending toward exclusion in so-called 'Hindu' legal texts and epics, but inclusion and transformation among Jainas and Buddhists (Deshpande 1999).

The mathematicians who wrote in Sanskrit, then, might have come from anywhere and their native language may be anyone's guess but they must have studied Sanskrit grammar. It is unlikely that they studied books. In India, no *paṇḍita* or traditional scholar does. An aspiring savant may have been taught by his father or must have had a teacher, who had his own guru, etc. in the oral succession of *guruparamparā*, "the lineage of teachers". All that knowledge must ultimately have come from one of the many existing and surviving Sanskrit grammatical works. Two questions arise: which one did it come from and what did it say?

Like other scholars, many mathematicians are likely to have studied the earliest and most famous Sanskrit grammar: that of Pāṇini of the fifth/fourth century BCE, or its later adaptations such as the *Kāśikā* of the seventh century CE or the Pāṇiniyan grammar of Bhaṭṭoji Dīkṣita of the seventeenth. Buddhists had their own grammars due to famous masters such as Candragomin of the fifth century CE, just as the Jainas had great grammarians from Devanandin (fifth c. CE) to Hemachandra (twelfth) and beyond (for more information on the Sanskrit grammarians see Staal 1972). All these works were inspired, directly or indirectly, by the Pāṇinian tradition. And all of them possessed not one but many zeroes.

What is zero in grammar or linguistics? Pāṇini had a technical term for it: *lopa*. He defined it as "something that does not appear"(*adarśanaṃ lopaḥ* 1.1.60). It is not a rare term in his grammar. Its "non-appearance" (*adarśanaṁ*) does not prevent it from occurring in forty-five out of four thousand rules if I counted them correctly as they were enumerated by Böhtlingk in his edition (1887, with many reprints: II: 271*). The actual number is higher since I have not taken account of Böhtlingk's uses of the expression *fgg* which indicates "and following".

Professor P.P. Divakaran, who commented on an earlier draft of this article, was intrigued by the definition of *lopa* as *adarśanaṃ* because Pāṇini certainly lived before Aśoka who presumably introduced writing in the third century BCE and: "I should think that a sound which is absent would be characterised by Pāṇini as unheard or unsounded rather than as unseen".

---

[1] The philologist's concept of *lectio difficilior*, "the more difficult reading," should be extended so as to be applicable to the oral transmission of compositions: the more difficult and unexpected of two readings or orally transmitted forms, viz., Āryabhaṭa, must be preferred to the normal, expected form Āryabhaṭṭa, since it is unlikely that it is based upon an error.

This is an apt observation but there is more to say. Some interpreters have indeed translated Pāṇini's *adarśanaṁ* as 'unseen' since the verbal root *dṛś* - certainly means "to see". I translated it differently as "something that does not appear" like others have done because that same root is widely used in the much wider meaning of seeing with one's mind. It includes perception, observation, appearance, knowing, etc., and is a common meaning in Sanskrit and similarly in other languages. In Indian philosophy, the six traditional systems are called *darśana* but their epistemology is not confined to seeing with one's eyes. In English we say: "I see what you mean". In later Sanskrit, *na dṛśyate* means: "it does not appear (that such-and-such is the case)". In English the same ambiguity applies to appear itself: "there appeared a large bird on the roof" suggests that the bird was actually seen but in "there appears to be much confusion about the PM of Thailand stepping down", nothing may have been peıceived by eyes or eyes only; it is a topic that people are talking about.

According to Renou's dictionary of grammatical terms in Sanskrit (Renou 1957 under *adarśana*, *lopa* and *lup-*), *adarśana* means *amuissement*, a technical term in French which expresses that a phoneme is dropped in pronunciation. Renou refers to Kātyāyana, grammarian who commented on Pāṇini and lived some two or three centuries later, probably during the reign of Aśoka (a fuller but somewhat opaque discussion occurs in Cardona: 1976, 1980, pages 267-273). That date could support an interpretation that refers to writing, but Kātyāyana himself does not see that way. Whether it was written or not, his statement is startling: *adarśana* means not seen, not heard, not pronounced, not perceived, absence or disappearance. It calls for a closer analysis and the entire subject seems to stand in need of a new and thorough revision but I think that, in the present context, we may safely conclude that *dṛś-* does not only mean "to see with one's eyes", that *adarśanam* does not only mean "unseen" and that "something that does not appear" is the best translation—for the time being.

Mathematicians studying Sanskrit in order to be able to compose works with all-India appeal, could not have missed the numerous zeroes that occur in Sanskrit grammars. Modern readers are able to understand them provided they know something of the morphology of words in Sanskrit. It is found in other Indo-European languages such as English; and in others. The examples that follow below under (1)-(3) occurred in three papers by Pandit: 1962, undated and 1990 (not seen), all of which, like Allen 1955, suffer from the complex constructs of outdated linguistic systems. No such defects disfigure Shefts 1961 who treated (2) and (3) and was reviewed in Staal 1963a (reprinted in 1988: 232-237).

Before we begin I must mention that Sanskrit does not use an explicit pronoun like English. The verbal form *khādati* does not mean: "eats" but "he eats". That "he" disappears when there is a subject as in *rāmo khādati* which means "*Rāma* eats" and not "*Rāma* he eats" (where the asterisk express ungrammatically) What is important in our context is that Sanskrit distinguishes like English between the

his name would have been Āryabhaṭṭa[1]); and neither need all brahmans know Sanskrit. *Ārya*, moreover, does not refer to "the three twice-born classes" as it probably did in the *Manusmṛti*, "The Law Code of Manu", 2.207 (Olivelle 2004: 242). The distinction between *ārya* and *anārya* has not been a racial or ethnic distinction. Madhav Deshpande has shown that it expresses claims to moral, social and spiritual status, tending toward exclusion in so-called 'Hindu' legal texts and epics, but inclusion and transformation among Jainas and Buddhists (Deshpande 1999).

The mathematicians who wrote in Sanskrit, then, might have come from anywhere and their native language may be anyone's guess but they must have studied Sanskrit grammar. It is unlikely that they studied books. In India, no *paṇḍita* or traditional scholar does. An aspiring savant may have been taught by his father or must have had a teacher, who had his own guru, etc. in the oral succession of *guruparamparā*, "the lineage of teachers". All that knowledge must ultimately have come from one of the many existing and surviving Sanskrit grammatical works. Two questions arise: which one did it come from and what did it say?

Like other scholars, many mathematicians are likely to have studied the earliest and most famous Sanskrit grammar: that of Pāṇini of the fifth/fourth century BCE, or its later adaptations such as the *Kāśikā* of the seventh century CE or the Pāṇiniyan grammar of Bhaṭṭoji Dīśkita of the seventeenth. Buddhists had their own grammars due to famous masters such as Candragomin of the fifth century CE, just as the Jainas had great grammarians from Devanandin (fifth c. CE) to Hemachandra (twelfth) and beyond (for more information on the Sanskrit grammarians see Staal 1972). All these works were inspired, directly or indirectly, by the Pāṇinian tradition. And all of them possessed not one but many zeroes.

What is zero in grammar or linguistics? Pāṇini had a technical term for it: *lopa*. He defined it as "something that does not appear"(*adarśanaṃ lopaḥ* 1.1.60). It is not a rare term in his grammar. Its "non-appearance" (*adarśanaṁ*) does not prevent it from occurring in forty-five out of four thousand rules if I counted them correctly as they were enumerated by Böhtlingk in his edition (1887, with many reprints: II: 271*). The actual number is higher since I have not taken account of Böhtlingk's uses of the expression *fgg* which indicates "and following".

Professor P.P. Divakaran, who commented on an earlier draft of this article, was intrigued by the definition of *lopa* as *adarśanaṃ* because Pāṇini certainly lived before Aśoka who presumably introduced writing in the third century BCE and: "I should think that a sound which is absent would be characterised by Pāṇini as unheard or unsounded rather than as unseen".

---

[1]The philologist's concept of *lectio difficilior*, "the more difficult reading," should be extended so as to be applicable to the oral transmission of compositions: the more difficult and unexpected of two readings or orally transmitted forms, viz., Āryabhaṭa, must be preferred to the normal, expected form Āryabhaṭṭa, since it is unlikely that it is based upon an error.

This is an apt observation but there is more to say. Some interpreters have indeed translated Pāṇini's *adarśanaṁ* as 'unseen' since the verbal root *dṛś*- certainly means "to see". I translated it differently as "something that does not appear" like others have done because that same root is widely used in the much wider meaning of seeing with one's mind. It includes perception, observation, appearance, knowing, etc., and is a common meaning in Sanskrit and similarly in other languages. In Indian philosophy, the six traditional systems are called *darśana* but their epistemology is not confined to seeing with one's eyes. In English we say: "I see what you mean". In later Sanskrit, *na dṛśyate* means: "it does not appear (that such-and-such is the case)". In English the same ambiguity applies to appear itself: "there appeared a large bird on the roof" suggests that the bird was actually seen but in "there appears to be much confusion about the PM of Thailand stepping down", nothing may have been peiceived by eyes or eyes only; it is a topic that people are talking about.

According to Renou's dictionary of grammatical terms in Sanskrit (Renou 1957 under *adarśana*, *lopa* and *lup-*), *adarśana* means *amuissement*, a technical term in French which expresses that a phoneme is dropped in pronunciation. Renou refers to Kātyāyana, grammarian who commented on Pāṇini and lived some two or three centuries later, probably during the reign of Aśoka (a fuller but somewhat opaque discussion occurs in Cardona: 1976, 1980, pages 267-273). That date could support an interpretation that refers to writing, but Kātyāyana himself does not see that way. Whether it was written or not, his statement is startling: *adarśana* means not seen, not heard, not pronounced, not perceived, absence or disappearance. It calls for a closer analysis and the entire subject seems to stand in need of a new and thorough revision but I think that, in the present context, we may safely conclude that *dṛś*- does not only mean "to see with one's eyes", that *adarśanam* does not only mean "unseen" and that "something that does not appear" is the best translation—for the time being.

Mathematicians studying Sanskrit in order to be able to compose works with all-India appeal, could not have missed the numerous zeroes that occur in Sanskrit grammars. Modern readers are able to understand them provided they know something of the morphology of words in Sanskrit. It is found in other Indo-European languages such as English; and in others. The examples that follow below under (1)-(3) occurred in three papers by Pandit: 1962, undated and 1990 (not seen), all of which, like Allen 1955, suffer from the complex constructs of outdated linguistic systems. No such defects disfigure Shefts 1961 who treated (2) and (3) and was reviewed in Staal 1963a (reprinted in 1988: 232-237).

Before we begin I must mention that Sanskrit does not use an explicit pronoun like English. The verbal form *khādati* does not mean: "eats" but "he eats". That "he" disappears when there is a subject as in *rāmo khādati* which means "*Rāma* eats" and not "*Rāma* he eats" (where the asterisk express ungrammatically) What is important in our context is that Sanskrit distinguishes like English between the

stem and the suffix or ending of a verb. From the stem *khād-* ("eat-") and the ending *-ti* ("-s") it forms:

khād-a-ti "(he) eat-s" (1)

What is *-a-* ? It is something in between which we may call an infix. I put hyphens between the three elements in the Sanskrit expression to distinguish the stem and the ending, which we find in both Sanskrit and English, and highlights the infix we only find in Sanskrit.

There are, however, various classes of verbs in Sanskrit. One of them has a verb with the same meaning but lacks the infix:

ad-ti "(he) eat-s" (2)

which becomes *atti* which is due to what is sometimes called "assimilation". Pāṇini's grammar is a list of rules (*sūtra*). (1) follows from a general rule. (2) seems to illustrate a special case. However, (1) and (2) express similar properties and the underlying rule is the same if we adopt a principle called *lāghava*, literally "lightness". (2) is now expressed as:

ad-0-*ti* "(he) eat-s" (3)

The symbol "0" which indicates absence of the infix is the *lopa* defined as "something that does not appear". Here we meet the grammatical or linguistic zero. It occurs not only in verbs but also in nouns and it should be obvious that there are many of them.

Pāṇini's rules are generally ordered and he could have reversed the order, i.e., start with (2) and derive (1) by inserting the infix. Such problems are discussed by commentators and other grammarians, including Joshi and Kiparsky 1979 and Kiparsky 1991 who have shown that "lightness" is not simply an abbreviation but expresses generalization. It explains the famous saying: "grammarians rejoice over the saving of half a syllable as over the birth of a son". My examples do no more than illustrate the thesis, that the many linguistic zeroes of Sanskrit grammar led mathematicians to muse about one or more mathematical zeroes.

Indic mathematicians did not only study Pāṇini to compose works with all-India appeal. They were a small group of people, not popular or prestigious like mathematicians in the modern world. In India there was only one "science of the sciences" (*śāstrāṇām śāstram*), the boundless (*anantapāram*), the summit of learning: grammar or *vyākaraṇa* which literally means "analysis". Mathematicians were flattered to be associated with such a prestigious tradition. In the modern world, the opposite holds. Grammar is not a popular subject and many scientists do not even know what "linguistics" means. It developed as a serious discipline only after the discovery of Sanskrit by William Jones, Charles Wilkins'

Sanskrit grammar and Franz Bopp's adoption of the methods of the Indian grammarians in his "Systems of Conjugation in Sanskrit compared with those in the Greek, Latin, Persian and Germanic Languages" of 1808 (see Thieme 1982/83, Staal 1989 and 1993, Chapters 4 and 5, and 1995: 109).

The histories of linguistics, logic and the theory of ritual in India and Europe were first compared in Staal 1988: 36-45. It presented graphs of developments that gave a rough idea of the ups and downs. Logic included mathematical logic, but I did not then, and would not now be able to include mathematics. "The Theory of Definition in Indian Logic" (1961, reprinted in 1988:90) referred to the occurrence of a kind of null-class presupposed in "modern" works such as the *Tarka-saṃgraha* of the seventeenth century AD. But in India, linguistics and logic were closely connected almost from the beginning.

Pāṇini distinguished different zeroes from each other by making use of a rich conceptual apparatus. He was aware of the fact that the language of his grammar was modeled in part after the language that was the object of his study: Sanskrit. It could lead to confusion unless the two were clearly distinguished. He therefore made a distinction between "rules" (*sūtra*) and "metarules" (*paribhāṣa*). Implicitly working with what we would now call a metalanguage, Pāṇini made explicit use of meta-linguistic markers which he called *it*. To distinguish the different zeroes from each other, he made use of the fact that *lopa* comes from a verbal root that starts with "l" and to which we shall return. The meta-linguistic markers always have an "l" that marks them as dealing with zeroes such as *luk*, *laṭ*, *liṭ* and *ślu*, each defined for particular classes or special cases. In modern transliterations, they are indicated by capitals (which Sanskrit has no means of distinguishing from small letters). The uses of metalanguage in Sanskrit grammar have been studied by Scharfe 1961 and Staal (1963b, 2003: 353-6); rules and metarules arestudied separately in Staal *forthcoming*.

English grammar does not use such meta-linguistic markers but it could do something similar. It may be illustrated with the help of a rough sketch of English noun pluralization (a formalized grammar of such a topic may look quite different and require a substantial book). We shall begin with a general rule, where *P* is the plural marker:

$$\text{noun} + P > \text{noun} + \text{suffix (e)s}$$

This is a context-sensitive rule in which > stands for "is replaced by"; + stands for concatenation; and parentheses express options that distinguish *dogs* from *witches*. The general rule as stated does not account for *fish* or *sheep* which require a zero-suffix.

My account, so far, applies to written English. It does not explain different pronunciations of the written *s*, which may sound like "*s*" or "*z*". If we try to account for pluralization in both written and spoken English we need a greater

variety of expressions. I shall not belabor the point but Pāṇini's way with metalinguistic zero-suffixes for special cases may be illustrated again for English by using subscripts as linguistic markers, for example:

$$\begin{aligned} man + 0_1 &> men \\ woman + 0_2 &> women \\ mouse + 0_3 &> mice. \end{aligned}$$

The spelling of English is idiosyncratic but that of Sanskrit, in that respect closer to Italian or Spanish, is rational. It is adopted by all Indic syllabaries in South and Southeast Asia and in the Roman transliteration adopted by Sanskrit scholars worldwide. That transliteration writes the *ou* of English *mouse* as *au* and the *i* of *mice* as *ai*. They are part of an extended system with similar sound correspondences in Sanskrit and Indo-European. Sanskrit derives from nouns such as *śiva* the adjective *Śaiva* which, in English, became "Shivaite". Similarly, the noun *rudra* produces the adjective *raudra* to which no English adjective corresponds.

4. Zeroes in the Śrauta Ritual

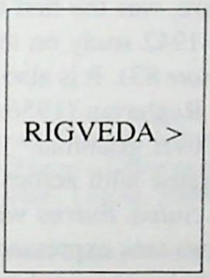
I shall end our discussion with the Vedic Śrauta ritual which belongs to the prehistory of zero as well as that of Sanskrit Grammar. Vedic ritual is, therefore, a parent as well as a grandparent of zero. All forms of Vedic ritual are concerned with recitations, chants and acts. Recitations come from the Rigveda, chants from the *Sāmaveda* and ritual acts are the chief concern of the Yajurveda. Vedic ritual became a science or theory (Staal, *forthcoming*) in one of its later forms which developed between roughly the tenth and seventh centuries BCE and became known as the Śrauta ritual. It was an oral tradition.

Śrauta ritual was the most creative of Vedic rituals. By that time, practice and theory had become highly developed disciplines and were closely connected with each other. That development did not take place among other higher animals which perform rituals but do not have theories because they do not possess language. Human ritualists are able to talk about ritual with each other, introduce modifications and add refinements. They started thinking ritually; but it did not happen often. Śrauta ritualists may have been the first. Japanese theologians influenced their rituals during recent centuries (Sharf 2003). Modern students have their own rituals but remain a motley crowd consisting of theologians, psychologists, representatives of disjointed human and social sciences while ritualization among the higher animals, again including the human, continues to be taken care of by life sciences on the borderline. Many of these specialists are innocent of theory. It explains an important difference between sections 3 and 4.

Sanskrit grammar deals with Sanskrit and influenced modern linguistics. The *Śrauta* ritual did not influence any modern theory of ritual because there is no such theory. It implies that we cannot adopt Needham's rule as formulated on page 3 above: we cannot do "the only thing we can do" because there is no yardstick. The reader will accordingly meet with unfamiliar concepts, methods and modes of analysis. Some are discussed in Chapters 7, 12 and 13 with their Source Notes in Staal 2008. But unless we are acquainted with one of the few surviving traditions of Śrauta ritual and/or some of the literature in Sanskrit and in modern languages about Vedic ritual, we shall be on our own.

Basic to any ritual performance are space and the four directions. The ritual arena consists of several sheds with thatched roofs that are temporarily constructed for each performance. Some of the most important *Śrauta* ceremonies are performed in a small space at the center that is called the Sadas. The word is derived from the Sanskrit verb *sad-* or "sit" which occurs in the contemporary term *upaniṣad*, "sitting close to" (the teacher).

The Sadas looks as follows with the north on top:



Vedas are recited in the four directions which the reciters themselves must also face. The above sketch makes use of two directions that are indicated in the figure by symbols we have used before but that now have a new meaning: > means "facing east" and < means "facing west". These directions raise a host of technical and theoretical problems, some of them discussed by Caland and Henry (1906-7: 232) and Keith (1914: I, 252 *note* 4). I shall mention two. The first is concerned with the directions only. The second combines directions with the verbal root from which *lopa* derives.

The first case is illustrated by the sketch. It depicts a change of directions that has just taken place. It does not involve the Rigveda which is recited inside the Sadas by an officiant who is already sitting there, facing and reciting east. I shall call him *R*. The Yajurveda is recited by a priest I shall call *Y*, but he has come from outside the Sadas and cannot easily enter. He has made several turns already and will eventually face west and face *R*: the auspicious result that is depicted here. Earlier, the two officiants did not only fail to face each other but *Y* sat with his back to *R* - a situation that is to be avoided at all costs. The entire episode illustrates how the rivalry between different factions may be resolved and overcome.

The stage is now set for the second case. A dialogue unfolds which is initiated by *R* who recites a proposal:

> "Let us both recite!" (*śoṃsāvo*)

Y responds from outside the Sadas with a touch of flattery. His verse consists of two halves:

> "Let us both recite, divine one!" (*śoṃsāvo daiva*)
>
> "Recite! Let us both rejoice!" (*śaṃsa madeva*)

Suppose *Y* were to omit the second half-verse. I have never witnessed it but it must have happened often because the two halves are very similar. The problem is addressed by the Yajurveda: "if the response after the half-verse were omitted (*lupyeta*) it would be like someone being left behind by others who are running ahead" (*Taittirīya Saṃhitā* 3.2.9.5). Here we have an instance of the verbal root *lup-* from which *lopa* is derived. The general meaning of the verb is "disappear" or "get lost".

Louis Renou, whom I mentioned before, was the first to draw attention to the numerous ritual uses of *lup-* in his 1941-1942 study on the connections between Sanskrit grammar and Vedic ritual (465, *note* 83). It is also Renou, "the most complete Sanskritist" as he was called by V. Raghavan (1956: 20), who first demonstrated the historical precedence of ritual over grammar.

I conclude that ancient India reverberated with zeroes, zero entities and zero events long before the geometry of the *Śulva Sūtra*s which are post-Buddhist. What may be called the prehistory of zero was expressed in early Vedic by *kha* which refers to cavities of various sorts and occurs in the *Upaniṣads* in the sense of "space". The *Śrauta Sūtras*, late Vedic but pre-Buddhist, used *lopa* to refer to omissions, disappearances and things that are lost. It is here that the origins of the mathematical concept of zero seem to lie. We do not know where it happened if it happened only once, but the most likely place would be the Kuru region north of modern Delhi though it may have been further east in Magadha, which overlaps with modern Bihar. The time must have been after 1,000 and before 600 BCE when the creative period of the Śrauta ritual was over. It is a long period with smudgy edges but there it is.

Reverting finally to arithmetics there are important questions that I have not so far considered. How was zero conceived as a member of the number series? Takao Hayashi has suggested that it may be related to additions such as 15 + 20 = 35 (in modern symbols) which presuppose 5 + 0 = 5 plausible enough but when did it happen? Or could zero be related to the recursive principle underlying the decimal number names (and hence their construction) which, according to P.P. Divakaran, was perfectly well understood in the early Rigveda? Plausible also, but would it imply that zero was conceived as the beginning of the infinite series of natural numbers, so that one would count 0, 1, 2, 3 etc.?

Let us return once more to our explanation of the origin of the mathematical zero in terms of the assignment of special meanings to Sanskrit terms such as *lopa*, *śunya* or *bindu*. We are fortunate to possess the records of such events. Similar words occur in other human languages but did not undergo a similar development as far as I know. In their Indic evolution, *lopa* was inspired by ritual but is that a necessary part of its prehistory? All we know is that it added the flesh of another empirical discipline to the bones of linguistics, the discipline that underlies our understanding of the development of language. As far as I can tell, thus far, two conditions must be satisfied before a concept of zero may arise: there needs to be a language as well as another formal structure in which that language is used to signify that something has disappeared or is lost.

Acknowledgments

It is with great pleasure that I like to express my gratitude to three colleagues who have assisted me substantially with this paper and of whom the first has already been mentioned: P. P. Divakaran, Sitabhra Sinha and Peter Vandemoortele.

References

[1] Allen, W.S. 1955. "Zero and Pāṇini", *Indian Linguistics* 16: 106-113

[2] Böhtlingk, Otto. 1887. Pāṇini's Grammatik, herausgegeben, überzetzt, erläutert und mit verschiedenen Indices versehen, Leipzig. Reprografischer Nachdruck, 1964, Hildesheim: Darmstadt.

[3] Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprachen*. Frankfurt.

[4] Bronkhorst, Johannes and Madhav M. Deshpande, eds. 1999. *Aryan and Non-Aryan in South Asia: Evidence, Interpretation and Ideology*. Proceedings of the International Seminar on Aryan and Non-Aryan in South Asia. Cambridge: Department of Sanskrit and Indian Studies. Harvard Oriental Studies. Opera Minora, Vol. 3.

[5] Caland. W. and Henry. V. 1906-7. L'*Agniṣṭoma. Description complète de la forme normale du sacrifice de Soma dans le culte védique*. Vols. I-II. Paris: Ernest Leroux.

[6] Cardona, George. 1976, 1980. *Pāṇini: A Survey of Research*. Mouton: The Hague and Delhi etc.: Motilal Banarsidass.

[7] Coedès, Georges. 1931. "A propos de l'origine des chiffres arabes," *Bulletin of the London School of Oriental and African Studies* 323-328.

[8] –, 1937-66. *Les inscriptions du Cambodge*. 8 Volumes. Hanoi and Paris.

[9] –, 1964. Les états hindouisés d'Indochine et d'Indonésie, Paris: E. de Boccard.

[10] Deshpande, Madhav M. 1999. "What to do with the *Anāryas*? Dharmic Discourse of Inclusion and Exclusion" in: Bronkhorst and Deshpande, eds.,107-127.

[11] Deshpande, Madhav M. and Saroja Bhate. 1991. *Pāṇinian Studies. Professor S.D. Joshi Felicitation Volume*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

[12] Divakaran, P.P. 2007. "The First Texbook of Calculus: *Yuktibhāṣā*", Indian Philosophy 35: 417-443.

[13] –, forthcoming. "Notes on *Yuktibhāṣā*: Recursive Methods in Indian Mathematics".

[14] Flood, Kevin, ed. 2003. *The Blackwell Companion to Hinduism*, UK: Blackwell.

[15] Hayashi, Takao. 2003. "Indian Mathematics", in: Flood, ed., 360-375.
[16] Joshi, S.D. and Paul Kiparsky. 1979. "Siddha and Asiddha in *Pāṇinian Phonology*" in Deshpande and Bhate: 223-250.
[17] Keith, Athur Berriedale. 1914. *The Veda of the Black Yajus School. Taittiriya Sanhita.* Vols. I-II. Cambridge, Mass: The Harvard University Press. Vol. 18
[18] Kiparsky, Paul. 1991. "Economy and the Construction of the Śivasūtras", in: Deshpande and Bhate.
[19] Minkowski, C. 2002. "Astronomers and their Reasons. Working Paper on Jyotiḥśāstra". *Journal of Indian Philosophy* 30: 495-514.
[20] –, 2008. "The Study of Jyotiḥśāstra and the Uses of Philosophy of Science", *Journal of Indian Philosophy* 36: 587-597.
[21] Nakayama, Shigeru and Sivin, Nathan. 1973. *Chinese Science. Exploration of an Ancient Tradition.* Cambridge, Mass.
[22] Nasr, Seyyed Hossein. 1968. *Science and Civilisation in Islam.* Cambridge, Mass. 1968.
[23] Needham, Joseph. 1959. *Science and Civilization in China* Vol.3: "Mathematics and the Sciences of the Heavens and the Earth". Cambridge: Cambridge University Press.
[24] –, 1976. *Science and Civilization in China* Vol.5: "Chemistry and Chemical Technology", Part III: Spagyrical Discovery and Invention: Historical Survey, from Cinnabar Elixirs to Synthetic Insulin. Cambridge: Cambridge University Press.
[25] Olivelle, Patrick. 2004. *The Law Code of Manu.* A New Translation Based on the Critical Edition. Oxford: University Press.
[26] Pandit, M. D. 1962. "Zero in Pāṇini", *Journal of the Maharaja Sayaji Rao University of Baroda* 11/1: 53-66.
[27] –, 1990. Zero in Pāṇini. Pune: Centre of Advanced Study in Sanskrit, University of Poona.
[28] –, undated. "Reflections on Pāṇinian Zero".
[29] Pingree, David. 1978. The *Yavanajātaka* of *Sphujidhvaja. Edited, translated and commented on.* Vols. I-II. Cambridge, Mass. and London, England: Harvard University Press. (Harvard Oriental Series, Vol. 48).
[30] –, 2003. "The logic of non-Western science: mathematical discoveries in medieval India". Daedalus 45-53.
[31] Raghavan, V. 1956. *Sanskrit and Allied Indological Studies in Europe.* Madras: University of Madras.
[32] Renou, Louis. 1941-42. "Les connexions entre le rituel et la grammaire en Sanskrit", *Journal asiatique* 233: 105-65. Republished in Staal 1972: 434-69.
[33] –, 1942, 1957. *Terminologie grammaticale du Sanskrit.* Paris: Champion.
[34] –, 1955-1969. *Etudes védiques et pa.ninéennes.* Vols. I - XVIII. Paris: E. de Boccard.
[35] –, 1964. *Hymnes á Agni. Deuxie'me partie.* In: Renou 1955-1969, 13: 1-163.
[36] Ruegg, D. Seyfort. 1959. *Contributions à l'Histoire de la Philosophie Linguistique Indienne.* Paris: E. de Boccard.
[37] –, 1978. "Mathematical and Linguistic Models in Indian Thought: The Case of Zero and *Śunyatā*," *Wiener Zeitschrift für die Kunde Südasiens* 12: 171-181.
[38] Scharfe, Hartmut. 1961. *Die Logik im Mahābhāṣya.* Berlin: Akademie-Verlag.
[39] de Solla Price, Derek J. 1971. "Joseph Needham and the Science of China" in: Nakayama and Sivin: 9-21.
[40] Sharf, Robert H. 2003. "Thinking Through Shingon Ritual", *Journal of the International Association of Buddhist Studies*, 26/1: 51-93.
[41] Shefts, Betty. 1961. *Grammatical Method in Paṇini: His Treatment of Sanskrit Present Stems.* American Oriental Series, Essay 1. New Haven: American Oriental Society.
[42] Staal, Frits. 1961. "The Theory of Definition in Indian Logic", *Journal of the American Oriental Society* 81: 122-126, reprinted in 1988: 88-92.
[43] –, 1963a. Review of Shefts 1961, *Language* 39: 483-488. Reprinted in Staal 1988: 232-237.

[44] –, 1963b. Review of Scharfe 1961, *Journal of the American Oriental Society* 83/2: 252-256. Reprinted in Staal 1988: 238-242.

[45] –, 1972. *A Reader on the Sanskrit Grammarians*, Cambridge, Mass and London, England: The *MIT* Press. *Studies in Linguistics*, Vol.I, edited by Samuel Jay Keyser.

[46] –, 1988. *Universals. Studies in Indian Logic and Linguistics*. Chicago and London: University of Chicago Press.

[47] –, 1989, 1993. *Rules without Meaning. Ritual, Mantras and the Human Sciences*. New York etc.: Peter Lang. Republished 1996 as *Ritual and Mantras: Rules without Meaning*. Delhi: Motilal Banarsidass.

[48] –, 1995. "The Sanskrit of Science", *Journal of Indian Philosophy* 23: 73-127.

[49] –, 2003. "The Science of Language" in: Flood, ed., 348-359.

[50] –, 2006. "Artificial Languages across Sciences and Civilizations", *Journal of Indian Philosophy* 34/1-2: 89-141.

[51] –, 2007. "Artificial Languages Between Innate Faculties", *Journal of Indian Philosophy* 35/5-6: 577-596.

[52] –, 2008. *Discovering the Vedas. Origins, Mantras, Rituals, Insights*, New Delhi: Penguin (first edition in March, revised edition in November).

[53] –, 2008a. "Two Whiffs of Air. A Critical Essay". *The Journal of Hindu Studies* (Oxford) 1: 130-147.

[54] –, *forthcoming*. "A Theory of Ritual". *Proceedings of the International Conference on Ritual Dynamics and the Science of Ritual*. University of Heidelberg: 29 September – 2 October, 2008.

[55] Thieme, Paul. 1982-83. "Meaning and form of the 'grammar' of Pāṇini". *Studien zur Indologie und Iranistik* 8/9: 3-34.

[56] Tropfke, Johannes. 1980. *Geschichte der Elementarmathematik*, 4. Auflage. Vol. 1: *Arithmetik und Algebra*. Vollständig neu bearbeitet von Kurt Vogel, KarinReich, Helmuth Gericke. Berlin: Walter de Gruyter.

Frits Staal
Professor Emeritus, University of California at Berkeley.

67 Moo 1, Ban Pong
Hang Dong, Chiang,
Mai Thailand 50230
E-mail: `fritsstaal@berkeley.edu`

Combinatorial Methods in Indian Music: *Pratyayas* in *Saṅgītaratnākara* of Sārṅgadeva

Raja Sridharan, R. Sridharan and M. D. Srinivas

1. Introduction

Six combinatorial tools (called *pratyayas*) have been in systematic use in India for the study of Sanskrit prosody (*Chandas-śāstra*) and these go back in time at least to Piṅgala (c. 300 BC?). Among these, three—*prastāra* (an enumeration rule for generating all the possible metrical patterns of a given class as a sequence of rows), *uddiṣṭa* (the process for finding, for any given metrical pattern, the corresponding row number in the *prastāra*) and *naṣṭa* (the converse of *uddiṣṭa*)—are found in Bharata's *Nāṭyaśāstra*, in the chapter where prosody is discussed. Incidentally, the problem of placing Piṅgala and Bharata in the chronology of time still remains an unsettled question. The notion of *pratyayas* was perhaps discussed in other ancient texts of music also. However, the first extant text on music where the *pratyayas* are systematically dealt with, both in connection with patterns of musical phrases (*tānas*) and patterns of musical rhythms (*tālas*), is *Saṅgītaratnākara* of Śārṅgadeva (c.1225 AD). Nārāyaṇa Paṇḍita in his *Gaṇitakaumudī* (1356 AD) deals with some of these questions in a more general context, though his theory does not cover the kind of *tāla-prastāra* considered by *Śārṅgadeva*.

Our aim in this article is to highlight the contributions of *Śārṅgadeva* and explain his work in a mathematical set up. We first discuss the sequential generation or enumeration of patterns of musical phrases, called *tāna-prastāra*. The method of generating these patterns, as discussed in the first chapter of *Saṅgītaratnākara*, is essentially a rule for generating sequentially the $n!$ permutations of n symbols. We note that the *prastāra*, and the *naṣṭa* and *uddiṣṭa* processes are all indeed encoded in a certain unique representation of any integer in terms of sums of factorials. We also explain how *Śārṅgadeva* employs a tabular figure, *khaṇḍa-meru*, to essentially go back and forth between any integer and its representation as a sum of factorials.

We then move on to discuss the *pratyayas* for patterns of musical rhythms, *tāla-prastāra*. This theory has been dealt with at length in the sixth chapter of

Saṅgītaratnākara. It is in fact a generalisation of the theory of *pratyayas* for moric metres or *mātrā-vṛttas*, where the short syllable (*laghu*) is taken to be of one *mātrā* (metrical time unit) and the long syllable (*guru*) is taken to be of two *mātrās*. *Saṅgītaratnākara* considers musical rhythmic patterns (*tālas*) made up of *druta* (of one time unit), *laghu*, *guru*, *pluta*, which are of 2, 4 and 6 durations respectively, of that unit of *druta*.

It should be noted that the notion of *pluta*, viewed as a prolated vowel, considered to be of the same duration as three *laghus*, appears already in ancient Sanskrit prosody. In fact the notion of *pluta* occurs in the *Ṛgveda* only three times, though much more frequently in other *Saṃhitās* and *Brāhmaṇas*. In classical Sanskrit, a pure consonant sound called a *vyañjana* (without any vowel attached) is said to be of half a *mātrā*. A *laghu* is said to be of one *mātrā*, *guru* of two *mātrās* and a *pluta* of three *mātrās* respectively. Clearly, as it is clumsy to handle fractions in music, time units of 1, 2, 4 and 6 were perhaps introduced and called *druta*, *laghu*, *guru* and *pluta* respectively.

Saṅgītaratnākara first presents a systematic method of enumerating all the *tālas* of a given time duration in a *prastāra*, and follows this up with a complete mathematical theory of *pratyayas* which is a generalisation of the corresponding theory for moric metres. An interesting feature of *tāla-prastāra* is that the total number of patterns (the *saṃkhyāṅka*), if laid out in a sequence, has a generating function which involves a polynomial of the sixth-degree. This is an analogue of the notion of *saṃkhyās* for the moric metres, which are given by the so-called Fibonacci numbers (a result which is already present in the seventh century Prākṛta text on prosody, *Vṛttajātisamuccaya* of Virahāṅka), which have as their generating function the inverse of the quadratic polynomial $(1 - x - x^2)$. In fact the sequence of *saṃkhyāṅkas* associated with *tāla-prastāras* satisfies a more complex (four term) recurrence relation and has the inverse of the sixth degree polynomial $(1 - x - x^2 - x^4 - x^6)$ as its generating function. Based on this generating function, we can compute the number of *tāla* patterns in the *prastāra*, which have a given number of *drutas*, *laghus*, *gurus* or *plutas*. This is an analogue of the *pratyaya* known as *ekadvayādi-lagakriyā* in Sanskrit prosody, and is discussed in *Saṅgītaratnākara* for the *tāla-prastāra* by introducing various tabular figures called *merus*, which are constructed on the basis of systematic recurrence relations which can be derived from the generating function mentioned above.

The discussion of *pratyayas* in prosody and music lead to the study of combinatorics related to three important ways of representing any non-negative integer, representations which are widely in use even today.[1] While discussing the *pratyayas* for *varṇa-vṛttas*, Piṅgala gave the procedure for finding the binary representation of integers. Much later, Nārāyaṇa Paṇḍita, in his *Gaṇitakaumudī*

[1] See for instance, A. S. Fraenkel, Systems of numeration, Amer. Math. Monthly 92 (1985), 105-114.

(c. 1356 AD), generalised the method of Piṅgala for an arbitrary radix. The discussion of *pratyayas* for *mātrā-vṛttas* led to the discovery of the so called Fibonacci representation of integers which expresses any integer uniquely as a suitable sum of Fibonacci numbers. In his discussion of *tāla-prastāra*, *Śārṅgadeva* introduced a generalization of the Fibonacci representation, where the Fibonacci numbers are replaced by the *saṃkhyāṅkas* of *tāla-prastāras*. Another form of generalization was considered later by Nārāyaṇa Paṇḍita. As we already noted, *Śārṅgadeva*, in his construction of *khaṇḍa-meru*, implicitly employed the so called factorial representation of integers in his discussion of the *pratyayas* for *tāna-prastāra*.

2. *Tāna-Prastāra*

Saṅgītaratnākara of Śārṅgadeva (c.1225) is one of the seminal texts of Indian music. It acts as a bridge between the ancient texts like *Nāṭya-śāstra*, *Bṛhaddeśi* and *Dattilam* and all the major later texts. As Śārṅgadeva states, his family originally came from Kashmir and his grandfather, father and he himself received patronage from king Siṅghaṇa (1210-1247) of the Yādava dynasty ruling from Devagiri in Maharashtra.

Chapter I of *Saṅgītaratnākara* is the *Svaragatādhyāya*. In this Chapter, the fourth section is entitled *Grāma-mūrcchanā-krama-tāna-prakaraṇa*. Of the 91 verses of this section, verses 61–70 deal with the *pratyayas* associated with *tāna-prastāra*. In the following discussion, we give the textual verses,[2] which discuss each of the *pratyayas*, along with a brief paraphrase[3] and necessary mathematical explanation.

A *tāna* is a sequence of the seven *svaras* (basic musical tones or notes) which we denote in their natural order by S, R, G, M, P, D, N. In this section, Śārṅgadeva considers the problem of enumeration of all the possible *tānas* where each *svara* appears only once.[4] For instance, if we consider the four *svaras* S, R, G, M,

---

[2]We have used the following edition of the work: *Saṅgītaratnākara* of Sārṅgadeva, edited with *Kalānidhi* of Kallinātha and *Saṅgītasudhākara* of Siṃhabhūpāla, by S. Subramanya Sastri, 2^{nd} Ed. revised by S. Sarada, Adyar Library, Madras: Vol. I, 1992; Vol. III, 1986. The excellent commentaries of Siṃhabhūpāla (c. 1350) and Kallinātha (c. 1420) present lucid explanations of the text along with examples.

[3]English translation of the verses on *Tānaprastāra*, in chapter I (*Svaragatādhyāya*) of the text, are available in: *Saṅgītaratnākara* of Sārṅgadeva edited with English translation by R. K. Shringy and Prem Lata Sharma, Vol. I, Motilal Banarsidass, Varanasi 1978. English translations of the verses on *Tālaprastāra*, in Chapter V (*Tālādhyāya*) of the text, are available in: Akella Mallikarjuna Sarma, *Tāla Prastāra of Niśśaṅka Śārṅgadeva's Saṅgītaratnākara*, Sannidhi Sangita Publications, Hyderabad 2001.

[4]Nārāyaṇa Paṇḍita in his *Gaṇitakaumudī* has generalised the theory of *tānaprastāra* to include *tānas* which involve repetition of *svaras* also.

there are 24 *tānas*, which represent the various possible permutations of them. Śārṅgadeva gives a systematic method of enumerating these permutations in a sequence. Before discussing his procedure for enumeration of the *tānas*, Śārṅgadeva first states the total number of number of *tānas*—this is the first *pratyaya* known as the *saṃkhyā* which is to count the number of *tānas* with a given number of *svaras*:

अङ्कानेकादिसप्तान्तानूर्ध्वमूर्ध्वं लिखेत् क्रमात्।
हते पूर्वेण पूर्वेण तेषु चाङ्के परे परे॥
एकस्वरादिसंख्या स्यात् क्रमेण प्रतिमूर्च्छनम्।

(सङ्गीतरत्नाकरः १.४.६१–२)

The above verses state that the total number of *tānas* with r *svaras* is given by the product $1.2\ldots r$, for $r = 1, 2, \ldots 7$. This is nothing but the statement that the total number of permutations that can be formed from r distinct *svaras* is $r!$.

2.1. The rule for the construction of the *Prastāra*

क्रमं न्यस्य स्वरः स्थाप्यः पूर्वः पूर्वः परादधः।
स चेदुपरि तत्पूर्वः पुरस्तूपरिवर्तिनः॥
मूलक्रमक्रमात् पृष्ठे शेषाः प्रस्तार ईदृशः।

(सङ्गीतरत्नाकरः १.४.६२-३)

The purport of the above verses is the following: All the *tānas* formed from a given set of distinct *svaras* can be enumerated in the form of an array in the following manner. In the first row the *svaras* are written in the ascending order. Successive rows in the *prastāra* are generated by the following rule: In each row of the *prastāra*, starting from the left, identify the first *svara* which has at least one lower *svara* to the left. Below that *svara*, place the highest of these lower *svaras*. Bring down all the *svaras* which are to the right, as they are, to the next row. To the left, place all the remaining *svaras* in the ascending order. This completes the next row of the *prastāra*. Continue till the last row is reached where all the *svaras* will be in the descending order.

In order to illustrate this rule, we may consider the case of four *svaras*: *S, R, G, M*. The first row of the *prastāra* is formed by placing them in the usual or ascending order $SRGM$. In order to form the next row, we scan the first row from the left and identify R as the first *svara* which has a lower *svara* S to the left. This S is placed below R. The G and M, to the right, are brought down as they are. The remaining *svara* R is placed to the left, thus completing the second row as $RSGM$. To get the third row, the second row is scanned from the left and G is identified as the first *svara*, which has a *svara* lower to the left.

Here there are actually two *svaras*, R, S, which are lower and are to the left of G, and the highest of them R is placed below G. The *svara* to the right M is brought down. The remaining *svaras*, S and G, are placed in the ascending order to the left, thus completing the third row as $SGRM$. And so on. The entire *prastāra* is shown in Table 1.

Śārṅgadeva's rule for the construction of the *prastāra* can be presented in a general context by considering the enumeration of the permutations of n distinct elements with a natural order, which we may assume to be $a_1, a_2, a_3, ..., a_n$. The *prastāra* of these is an $n! \times n$ array of n symbols in $n!$ rows, which will be denoted by $[a_1, a_2, a_3, ...a_n]$. We will also denote by $[a_i]_r$ the $1 \times r$ column vector with all the entries given by a_i.

1	S	R	G	M		a_1	a_2	a_3	a_4
2	R	S	G	M		a_2	a_1	a_3	a_4
3	S	G	R	M		a_1	a_3	a_2	a_4
4	G	S	R	M		a_3	a_1	a_2	a_4
5	R	G	S	M		a_2	a_3	a_1	a_4
6	G	R	S	M		a_3	a_2	a_1	a_4
7	S	R	M	G		a_1	a_2	a_4	a_3
8	R	S	M	G		a_2	a_1	a_4	a_3
9	S	M	R	G		a_1	a_4	a_2	a_3
10	M	S	R	G		a_4	a_1	a_2	a_3
11	R	M	S	G		a_2	a_4	a_1	a_3
12	M	R	S	G		a_4	a_2	a_1	a_3
13	S	G	M	R		a_1	a_3	a_4	a_2
14	G	S	M	R		a_3	a_1	a_4	a_2
15	S	M	G	R		a_1	a_4	a_3	a_2
16	M	S	G	R		a_4	a_1	a_3	a_2
17	G	M	S	R		a_3	a_4	a_1	a_2
18	M	G	S	R		a_4	a_3	a_1	a_2
19	R	G	M	S		a_2	a_3	a_4	a_1
20	G	R	M	S		a_3	a_2	a_4	a_1
21	R	M	G	S		a_2	a_4	a_3	a_1
22	M	R	G	S		a_4	a_2	a_3	a_1
23	G	M	R	S		a_3	a_4	a_2	a_1
24	M	G	R	S		a_4	a_3	a_2	a_1

TABLE 1. *Tāna-prastāra* of SRGM.

We explain the rule of enumeration by an inductive procedure. To start with, the *prastāra* $[a_1]$ of one element is simply the 1×1 column $[a_1]_1$. The *prastāra*

$[a_1, a_2]$ of two elements is the 2×2 array

$$\begin{matrix} a_1 \, a_2 \\ a_2 \, a_1 \end{matrix}$$

which can also be written as

$$\begin{matrix} [a_1] \, [a_2]_1 \\ [a_2] \, [a_1]_1 \end{matrix}$$

The *prastāra* $[a_1, a_2, a_3]$ of three elements is a 6×3 array, which (in our notation) has the form

$$\begin{matrix} [a_1, a_2][a_3]_2 \\ [a_1, a_3][a_2]_2 \\ [a_2, a_3][a_1]_2 \end{matrix}$$

Having obtained the *prastāra* of three elements $[a_1, a_2, a_3]$, we can see that the *prastāra* of four elements $[a_1, a_2, a_3, a_4]$, explicitly shown in the preceding Table 1, has the structure

$$\begin{matrix} [a_1, a_2, a_3] \, [a_4]_6 \\ [a_1, a_2, a_4] \, [a_3]_6 \\ [a_1, a_3, a_4] \, [a_2]_6 \\ [a_2, a_3, a_4] \, [a_1]_6 \end{matrix}$$

By induction, the *prastāra* $[a_1, a_2, ..., a_n]$ of n elements can be shown to be the $n! \times n$ array given by

$$\begin{matrix} [a_1, a_2, ..., a_{n-2}, a_{n-1}] \, [a_n]_{(n-1)!} \\ [a_1, a_2, ..., a_{n-2}, a_n] \, [a_{n-1}]_{(n-1)!} \\ \vdots \qquad\qquad \vdots \\ [a_2, a_3, ..., a_{n-1}, a_n] \, [a_1]_{(n-1)!} \end{matrix}$$

Here, we first note that the n-th element changes for the first time (from a_n to a_{n-1}) after $(n-1)!$ rows. The permutation with row-number $(n-1)!$ has the form $(a_{n-1}, a_{n-2} ..., a_1, a_n)$ in which the first $n-1$ elements appear in the reverse order, followed by a_n. In this permutation, only the last two elements are in the ascending order and, following Śārṅgadeva's rule for *prastāra*, the succeeding row (with row-number $(n-1)!+1$) is constructed by placing the largest of these elements (in this case a_{n-1}) below a_n and placing the rest of the elements in the ascending order

at the left. This is precisely the permutation $(a_1, a_2, ..., a_{n-2}, a_n, a_{n-1})$ which appears in the above general form at the row numbered $(n-1)! + 1$. By induction, it can be seen that the array given above is indeed the *prastāra* of n elements according to the construction of Śārṅgadeva.

We may also note that the above *prastāra* enumerates all the permutations of n distinct elements in an order which is different from the so-called lexicographic order, which is perhaps more commonly used in modern combinatorics. Śārṅgadeva's rule actually enumerates the permutations in the what is sometimes referred to as the colex order—the mirror image of lexicographic order in the reverse.

2.2. A "factorial representation" of integers

After discussing the *prastāra* or the enumeration process, Śārṅgadeva discusses the *pratyayas*, *uddiṣṭa* and *naṣṭa*. *Uddiṣṭa* is the process by which, given a *tāna* or permutation of the *svaras*, one finds the number of the row in which it appears in the *prastāra* (or enumeration), without going through the process of generating the entire *prastāra*. *Naṣṭa* is the converse process by which, given the number of the row, the corresponding *tāna* or the permutation is determined. The processes of *uddiṣṭa* and *naṣṭa* are discussed by Śārṅgadeva in terms of what is called the *khaṇḍa-meru*, and are essentially based upon a certain representation of any integer uniquely as a sum of factorials.

In this section we shall outline some mathematical aspects of this representation, which is a variant of what is widely known as the factorial representation of integers. We shall then consider how Śārṅgadeva's construction of *prastāra* and the *uddiṣṭa* and *naṣṭa* processes are encoded in this representation. We begin with the following:

Proposition 1: Every integer $1 \leq m \leq n!$ can be uniquely represented in the form

$$m = d_0 0! + d_1 1! + d_2 2! + ... + d_{n-1}(n-1)!, \tag{1}$$

where d_i are integers such that $d_0 = 1$ and $0 \leq d_i \leq i$, for $i = 1, 2, \ldots, n-1$.

Proof: We prove the existence of such a representation by actually giving an explicit procedure by which every integer $1 \leq m \leq n!$ can be represented in the form (1) given above. First, d_{n-1} is obtained as the quotient in the division of m by $(n-1)!$, where we stipulate further that, if m is divisible by $(n-1)!$, then the quotient is reduced by 1 and $(n-1)!$ is taken as the remainder. That is, d_{n-1} and r_{n-1} are given uniquely by

$$m = d_{n-1}(n-1)! + r_{n-1} \tag{2}$$

where, $0 \leq d_{n-1} \leq n-1$ and $0 < r_{n-1} \leq (n-1)!$. Next, we find d_{n-2} and r_{n-2} similarly by dividing r_{n-1} by $(n-2)!$. And so on. When we finally reach r_2, since $0 < r_2 \leq 2$, we set $d_1 = 1$ if $r_2 = 2$ and $d_1 = 0$ if $r_2 = 1$. Then, it follows that $d_0 = 1$, and we have thus a systematic procedure by which every integer $1 \leq m \leq n!$ can be represented in the form (1) given above. In particular, for any $r \leq n$ we have a fascinating relation (with $d_0 = 1, d_i = i$, for all $1 \leq i \leq r-1$),

$$r! = 1.0! + 1.1! + 2.2! + \ldots + (r-1)(r-1)! \tag{3}$$

The uniqueness of the representation (1) can be shown by induction on m. We note that since the uniqueness is clear for $n = 1, 2$, we may certainly start the induction. Assume that $m \geq 2$ and let

$$\begin{aligned} m &= 1.0! + d_1 1! + d_2 2! + \ldots + d_{n-1}(n-1)! \\ &= 1.0! + d_1'! + d_2' 2! + \ldots + d_{n-1}'(n-1)! \end{aligned}$$

If $d_{n-1} = d_{n-1}' \neq 0$, then

$$\begin{aligned} m - d_{n-1}(n-1)! &= 1.0! + d_1 1! + d_2 2! + \ldots + d_{n-2}(n-2)! \\ &= 1.0! + d_1' 1! + d_2' 2! + \ldots + d_{n-2}'(n-2)!, \end{aligned}$$

so that (by induction on m) it follows that $d_i = d_i'$ for $1 \leq i \leq (n-2)$ also. If $d_{n-1} \neq d_{n-1}'$, then we can assume without loss of generality that $d_{n-1} > d_{n-1}'$ and we have

$$\begin{aligned} m - d_{n-1}'(n-1)! &= 1.0! + d_1 1! + d_2 2! + \ldots + d_{n-2}(n-2)! \\ &\quad + (d_{n-1} - d_{n-1}')(n-1)! \\ &= 1.0! + d_1'! + d_2' 2! + \ldots + d_{n-2}'(n-2)! \end{aligned}$$

If $d_{n-1}' > 0$, then $m - d_{n-1}'(n-1)! < m$, and we are through by induction. On the other hand, if $d_{n-1}' = 0$, then since $d_{n-1} > 0$ the first of the above two equations yields $m > (n-1)!$, while the second (in view of (3)) yields $m \leq (n-1)!$, a contradiction. This proves the proposition.

Remark: The above representation (1) is in fact a variant of what is generally referred to as the factorial representation of integers, where every integer $1 \leq m \leq n!$, has a unique representation of the form

$$m = f_1 1! + f_2 2! + \ldots + f_{n-1}(n-1)! + f_n n!, \tag{4}$$

with $0 \leq f_i \leq i$, for $i = 1, 2, \ldots n$. If $m = n!$, we have $f_i = 0$, for $i < n$ and $f_n = 1$; in other words, the above representation (3) becomes trivial when $m = n!$. On the other hand, the factorial representation (1), given by the above

proposition, reduces to the following relation (with $d_i = i$, for all i),

$$n! = 1.0! + 1.1! + 2.2! + \ldots + (n-1)(n-1)!,$$

which has been mentioned already.

Corollary: For any integer $1 \leq m \leq n!$, the assignment $m \rightarrow (d_1, d_2, \ldots d_{n-1})$, given by equation (1), is a bijection of the set of natural numbers $\leq n!$, with the set of all $(n-1)$-tuples of integers $\{(d_1, d_2, \ldots, d_{n-1}) | 0 \leq d_i \leq i, i = 1, 2, ..., n-1\}$.

2.3. *Uddiṣṭa* and *naṣṭa* processes and the factorial representation

We shall use the factorial representation discussed above to understand the processes of *uddiṣṭa* and *naṣṭa* as also the method of construction of the *prastāra*. We shall first show that there is a one to one correspondence between the set $(a_{i_1}, a_{i_2}, \ldots, a_{i_n})$ of all permutations of n distinct symbols and the coefficients $\{d_0 = 1, d_1, d_2, \ldots, d_{n-1} | 1 \leq d_i \leq i,\ i = 1, 2, \ldots, n-1)\}$ which arise in the factorial representation of numbers $1 \leq m \leq n!$. As an illustration of this correspondence, we present in Table 2 the *prastāra* $[a_1, a_2, a_3, a_4]$ of four distinct elements along with the coefficients $d_0 = 1, d_1, d_2, d_3$ of the factorial representation for each row-number m for $1 \leq m \leq 24$.

Given any permutation $(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)})$, the corresponding set of coefficients $d_1, d_2, \ldots, d_{n-1}$ are determined sequentially, starting first with d_{n-1}, by carefully following the inductive process that we described for generating the *prastāra*. As we mentioned earlier, the *prastāra* for n elements $[a_1, a_2, ..., a_{n-1}, a_n]$ can be written down using the *prastāra* for $n-1$ elements in the following manner:

$$\begin{array}{c} [a_1, a_2, ..., a_{n-2}, a_{n-1}]\ [a_n]_{(n-1)!} \\ [a_1, a_2, ..., a_{n-2}, a_n]\ [a_{n-1}]_{(n-1)!} \\ \vdots \qquad\qquad \vdots \\ [a_2, a_3, ..., a_{n-1}, a_n]\ [a_1]_{(n-1)!} \end{array}$$

Each of the blocks in the above *prastāra* has $(n-1)!$ rows. Thus, the last element in the first $(n-1)!$ rows is a_n and this corresponds to $d_{n-1} = 0$ in the factorial representation of the corresponding row-number. In the next $(n-1)!$ rows, the last element is a_{n-1} and this corresponds to $d_{n-1} = 1$. And so on. Thus we see that if the last element of the given row is $a_{\sigma(n)}$ where σ is a permutation of $1, 2, ..., n$, then the corresponding row-number will have a factorial representation with $d_{n-1} = n - \sigma(n)$. Then, we remove the element $a_{\sigma(n)}$ and consider the *prastāra* of the remaining $n-1$ elements and determine the coefficient d_{n-2} in

the same way as we found d_{n-1} by considering the last element $a_{\sigma(n-1)}$ of the permutation $(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n-1)})$. For this purpose, let $\sigma(n-1)$, be the j-th element when we arrange $\sigma(1), \sigma(2), \ldots, \sigma(n-1)$ in the increasing order. Then, $d_{n-2} = n - 1 - j$. Again, let $\sigma(n-2)$, be the k-th element when we arrange $\sigma(1), \sigma(2), \ldots, \sigma(n-2)$ in the increasing order. Then, $d_{n-3} = n - 2 - k$. We continue this procedure obtaining d_i up to $i = 1$, and then set $d_0 = 1$.

m					d_0	d_1	d_2	d_3
1	a_1	a_2	a_3	a_4	1	0	0	0
2	a_2	a_1	a_3	a_4	1	1	0	0
3	a_1	a_3	a_2	a_4	1	0	1	0
4	a_3	a_1	a_2	a_4	1	1	1	0
5	a_2	a_3	a_1	a_4	1	0	2	0
6	a_3	a_2	a_1	a_4	1	1	2	0
7	a_1	a_2	a_4	a_3	1	0	0	1
8	a_2	a_1	a_4	a_3	1	1	0	1
9	a_1	a_4	a_2	a_3	1	0	1	1
10	a_4	a_1	a_2	a_3	1	1	1	1
11	a_2	a_4	a_1	a_3	1	0	2	1
12	a_4	a_2	a_1	a_3	1	1	2	1
13	a_1	a_3	a_4	a_2	1	0	0	2
14	a_3	a_1	a_4	a_2	1	1	0	2
15	a_1	a_4	a_3	a_2	1	0	1	2
16	a_4	a_1	a_3	a_2	1	1	1	2
17	a_3	a_4	a_1	a_2	1	0	2	2
18	a_4	a_3	a_1	a_2	1	1	2	2
19	a_2	a_3	a_4	a_1	1	0	0	3
20	a_3	a_2	a_4	a_1	1	1	0	3
21	a_2	a_4	a_3	a_1	1	0	1	3
22	a_4	a_2	a_3	a_1	1	1	1	3
23	a_3	a_4	a_2	a_1	1	0	2	3
24	a_4	a_3	a_2	a_1	1	1	2	3

TABLE 2. *Tāna-prastāra* and Factorial Representation.

Having determined the coefficients $d_0, d_1, \ldots, d_{n-1}$ associated with the given permutation $(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)})$, the corresponding row-number m, in the *prastāra* is given by

$$m = d_0 0! + d_1 1! + d_2 2! + \ldots + d_{n-1}(n-1)!$$

This procedure for determining the row number of a given permutation is called (as in prosody) *uddiṣṭa*. Śārṅgadeva follows a similar method for the *uddiṣṭa* process which we shall describe later.

As an example of the above procedure, we may, for instance, compute the row number of the permutation (a_3, a_2, a_4, a_1) in the *prastāra* $[a_1, a_2, a_3, a_4]$. Here $\sigma(1) = 3$, $\sigma(2) = 2$, $\sigma(3) = 4$ and $\sigma(4) = 1$. According to the procedure outlined above, $d_3 = 4 - \sigma(4) = 4 - 1 = 3$. The element $a_{\sigma(3)} = a_4$, occurs as the third element of a_2, a_3, a_4 and hence $d_2 = 3 - 3 = 0$. The element $a_{\sigma(2)} = a_2$, occurs as the first element of a_2, a_3 so that $d_1 = 2 - 1 = 1$. Hence the row number of (a_3, a_2, a_4, a_1) is $1.0! + 1.1! + 0.2! + 3.3! = 20$.

We now indicate the converse process of *naṣṭa* (or finding the "lost" row), which gives a procedure to find, for any integer $m \leq n!$, the m-th row of the *prastāra* $[a_1, a_2, ..., a_n]$. We begin with the factorial representation of m:

$$m = d_0 0! + d_1 1! + d_2 2! + ... + d_{n-1}(n-1)!$$

We set $\sigma(n) = n - d_{n-1}$ and take $a_{\sigma(n)}$ as the last element of the row. We remove $a_{\sigma(n)}$ from the set $a_1, a_2, ..., a_n$ and arrange the rest of the elements in the ascending order. If $j = n - 1 - d_{n-2}$, then we let the j-th element of this sequence be the penultimate element $a_{\sigma(n-1)}$ of the m-th row. We next omit $a_{\sigma(n)}$ and $a_{\sigma(n-1)}$ and arrange the rest of the elements of the set $a_1, a_2, ..., a_n$ in the ascending order. If $k = n - 2 - d_{n-3}$, then the k-th element in this sequence will be the next element $a_{\sigma(n-2)}$ in the m-th row. We continue this procedure till we use up all d_i, $1 \leq i \leq (n-1)$, and define $a_{\sigma(1)}$ to be the remaining symbol left in $a_1, a_2, ..., a_n$. Thus the permutation in the m-th row is determined to be $(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$. This is the *naṣṭa* process. Śārṅgadeva also follows essentially the same method as we shall see later.

To illustrate the above procedure, we shall find the 15th row of the *prastāra* $[a_1, a_2, a_3, a_4]$. We have $15 = 1.0! + 0.1! + 1.2! + 2.3!$. The rule above shows that $\sigma(4) = 4 - 2 = 2$. The last element in the row is thus a_2. The remaining elements in the increasing order of the indices are a_1, a_3, a_4. Since, $\sigma(3) = 3 - 1 = 2$, we take the second element amongst them, namely a_3, as the penultimate element of the row. Now we omit a_3 also and are left with a_1, a_4. Since, $\sigma(2) = 2 - 0 = 2$, the next element of the row will be a_4. The remaining symbol is a_1. Thus the 15th row of the *prastāra* is determined to be (a_1, a_4, a_3, a_2).

The factorial representation also sheds considerable light on the method of constructing the *prastāra*. To see this, we take one more look at the way the *prastāra* of n elements is constructed from that of $n - 1$ elements.

$$\begin{array}{c}
[a_1, a_2, ..., a_{n-2}, a_{n-1}][a_n]_{(n-1)!} \\
[a_1, a_2, ..., a_{n-2}, a_n][a_{n-1}]_{(n-1)!} \\
\vdots \qquad\qquad \vdots \\
[a_2, a_3, ..., a_{n-1}, a_n][a_1]_{(n-1)!}
\end{array}$$

Here, we note that the n-th element changes for the first time (from a_n to a_{n-1}) after $(n-1)!$ rows. And, at the same time, the co-efficient d_{n-1} in the factorial representation changes from 0 to 1. In other words, for numbers $1 \leq m \leq (n-1)!$, $d_{n-1} = 0$; and we have $d_{n-1} = 1$ for the first time when $m = (n-1)! + 1$. Now, the number $(n-1)!$ has the factorial representation

$$(n-1)! = 1.0! + 1.1! + 2.2! + ... + (n-2)(n-2)! + 0.(n-1)!$$

Here, we have (apart from the usual $d_0 = 1$), $d_i = i$ for $1 \leq i \leq n-2$, and $d_{n-1} = 0$. When we go to the next row, the representation is changed to

$$(n-1)! + 1 = 1.0! + 0.1! + 0.2! + ... + 0.(n-2)! + 1.(n-1)!$$

That is, in passing from $(n-1)!$ to $(n-1)! + 1$, the coefficient d_{n-1} increases by 1, and the coefficients d_i change to 0 for $1 \leq i \leq n-2$. By using induction on n, we can easily establish the following more general result:

Proposition 2: Let m be an integer such that $1 \leq m \leq n!$, with factorial representation

$$m = 1.0! + d_1 1! + d_2 2! + ... + d_{n-1}(n-1)!,$$

with $0 \leq d_i \leq i$, for $i = 1, 2, ...n-1$. If $d_k (k > 1)$ is the first coefficient in the above factorial representation such that $d_k < k$, then the factorial representation for $(m+1)$ is given by[5]

$$m + 1 = 1.0! + d'_1 1! + d'_2 2! + ... + d'_{n-1}(n-1)!,$$

where $d'_i = d_i - i = 0$ for $i < k$, $d'_k = d_k + 1$, and $d'_i = d_i$ for $i > k$.

This proposition gives the precise relation between the factorial representation and the rule of Śārṅgadeva for generating successive rows in the *prastāra*, discussed earlier.

2.4. *Uddiṣṭa* and *Naṣṭa* processes as discussed by Śārṅgadeva

2.4.1. *Khaṇḍa-meru*

We now present the *uddiṣṭa* and *naṣṭa* processes as discussed by Śārṅgadeva and show how they are indeed based on the factorial representation of integers

[5]If it happens that $d_i = i$ for all $1 \leq i \leq (n-1)$, then from (4) we have $m = n!$, and we are at the end of the *prastāra*.

discussed above. Śārṅgadeva bases his discussion of the *uddiṣṭa* and *naṣṭa* processes through a tabular figure referred to as the *khaṇḍa-meru*, which is essentially a device to read off the co-efficients d_i which arise in the factorial representation of any given number and vice versa. The *khaṇḍa-meru* is described in the following verses:

सप्तादेकान्तकोष्ठानामधोऽधः सप्तपङ्क्तयः ॥
तास्वाद्यायाद्यकोष्ठे लिखेदेकं परेषु खम्।
वेद्यतानस्वरमितान् न्यसेत् तेष्वेव लोष्टकान्॥
प्राक्पङ्क्त्यन्त्याङ्कसंयोगमूर्ध्वाधः स्थितपङ्क्तिषु।
शून्यादधो लिखेदेकं तं चाधोऽधः स्वकोष्ठकान् ॥
कोष्ठसंख्यागुणं न्यस्येत् खण्डमेरुरयं मतः।

(सङ्गीतरत्नाकरः १.४.६३-६)

The tabular figure *khaṇḍa-meru* consists of seven columns (as we are dealing with the seven *svaras* in *tāna-prastāra*) with the number of entries increasing from one to seven, from left to right. In the first row, place 1 in the first column (from the left) followed by $0's$ in the other columns. In the second row, starting from the second column, place the total number (or *sāṃkhyā*) of *tānas* of 1, 2, 3, etc., *svaras*. Thus the entries in the second row are the factorials, 1, 2, 6, 24, 120 and 720. In the succeeding rows, place twice, thrice etc. (that is successive multiples) of the factorials, starting from a later column at each stage. Thus, the first column of the *meru* consists of just $0! = 1$, and the n-th column, for $n > 1$, is made up of the multiples $i.(n-1)!$, with $0 \leq i \leq n-1$, as shown in Table 3.

S	R	G	M	P	D	N
1	0	0	0	0	0	0
	1	2	6	24	120	720
		4	12	48	240	1440
			18	72	360	2160
				96	480	2880
					600	3600
						4320

TABLE 3. *Khaṇḍa-meru.*

Thus, in the tabular figure of *khaṇḍa-meru*, Śārṅgadeva has a table of multiples of all the factorials. He uses this effectively to find out the number of the row in a *prastāra* that corresponds to a given *tāna* (the *uddiṣṭa* process), and conversely (the *naṣṭa* process), by precisely making use of the factorial representation that we discussed earlier. This will be clear from what follows.

2.4.2. *Uddiṣṭa Process*

Śārṅgadeva explains the *uddiṣṭa* process as follows:

स्वरान् मूलक्रमस्यान्त्यात् पूर्वं यावतिथः स्वरः ॥
उद्दिष्टान्त्यस्तावतिथे कोष्ठेऽधो लोष्टकं क्षिपेत्।
लोष्टचालनमन्त्यात् स्यात्त्यक्त्वा लब्धं क्रमो भवेत् ॥
लोष्टाक्रान्ताङ्कसंयोगादुद्दिष्टस्य मितिर्भवेत्॥

(सङ्गीतरत्नाकरः १.४.६६-८)

Given a *tāna* (of n *svaras*), note the rank of the last *svara* counting the *svaras* from the end in the natural order (i.e., counting the *svaras* in the reverse of the natural order). Mark the corresponding entry in the last or the n-th column of the *khaṇḍa-meru.* Note the rank of the next *svara* (in the reverse of natural order) among the remaining *svaras.* Mark the corresponding entry in the next or the $(n-1)$-th column. And so on. The *uddiṣṭa* or the rank-number of the given *tāna* is the sum of the all the marked entries.

S	R	G	M	P	D	N
1	0	0	0	0	0	0
	1	**2**	**6**	24	120	720
		4	12	48	240	1440
			18	72	360	2160
				96	480	2880
					600	3600
						4320

TABLE 4. *Khaṇḍa-meru* for *uddiṣṭa.*

Example: Rank of the *tāna* $MSRG$.

- 4-th column: G is second of S, R, G, M in the reverse order, so mark the second entry 6.
- 3-rd column: R is second among (the remaining) S, R, G in the reverse order, so mark 2.
- 2-nd column: S is the second among S, R in the reverse order, so mark 1.
- 1-st column: M is left, and we mark the entry 1.

Rank-number $= 1 + 1 + 2 + 6 = 10$.

Clearly, the process of *uddiṣṭa*, as set forth above by Śārṅgadeva, involves first working out the co-efficients d_i which correspond to the given *tāna* or permutation. Then the *khaṇḍa-meru* is employed to find out the row-number associated with these co-efficients. In the case of the given *tāna*, $MSRG$, the coefficients turn out to be $d_3 = d_2 = d_1 = 1$. Then the corresponding row-number is found from the *khaṇḍa-meru* to be

$$1.0! + 1.1! + 1.2! + 1.3! = 1 + 1 + 2 + 6 = 10.$$

2.4.3. Naṣṭa process

Śārṅgadeva explains the converse or the *naṣṭa* process as follows:

येरङ्कैर्नष्टसंख्या स्यान्मौलैकाङ्कसमन्वितैः ॥
तेषु लोष्टं क्षिपेन्मूले लोष्टस्थानमितं भवेत् ।
नष्टतानस्वरस्थानं ततो यावतिथे पदे ॥
अधःक्रमादस्तिलोष्टः स्वरस्तावतिथो भवेत्।
क्रमान्तिमस्वरात्पूर्वो लब्ध्यत्यागादि पूर्ववत्॥

(सङ्गीतरत्नाकरः १.४.६८-७०)

To find the *tāna* (of n *svaras*) corresponding to a given rank-number or *naṣṭa-saṃkhyā*, mark that entry which is the greatest number strictly less than the rank-number in the n-th column of the *khaṇḍa-meru*. Subtract that entry from the rank number and mark the entry, which is the greatest number strictly less than the resulting number, in the next or the $(n-1)$-th column. And so on. The position of the marked entry in the last column gives the rank of the last *svara* of the *tāna* in the reverse natural order. The position of the marked entry in the next column gives the rank of the last but one *svara*, amongst the remaining *svaras*, in the reverse natural order. And so on.

Example: To find the 18^{th} *tana* in the *prastāra* of $SRGM$

- In the fourth column of the *khaṇḍa-meru*, the greatest number less than 18 is 12, which occurs on the third row. The fourth *svara* of the *tāna* is thus the third among S, R, G, M in reverse order, namely R.
- Now, $18 - 12 = 6$. The greatest number less than this in the third column is 4, which occurs in the third row. The third *svara* of the *tāna* is the third among S, G, M in reverse order, S.
- Now, $6 - 4 = 2$. The greatest number less than this in the second column is 1, occurring in the second row. Second *svara* of the *tāna* is second among G, M in reverse order, G.
- The first *svara* of the *tāna* is the one left, M.

S	R	G	M	P	D	N
1	0	0	0	0	0	0
	1	2	6	24	120	720
		4	**12**	48	240	1440
			18	72	360	2160
				96	480	2880
					600	3600
						4320

TABLE 5. *Khaṇḍa-meru* for *naṣṭa*.

The 18^{th} *tāna* is therefore $MGSR$.

Clearly, the process of *naṣṭa*, as set forth above by Śārṅgadeva, involves first working out the factorial representation of the given number. In the above example, the number 18 was decomposed as

$$18 = 1 + 1 + 4 + 12 = 1.0! + 1.1! + 2.2! + 2.3!$$

Here, the *khaṇḍa-meru* is used to straight away read off the coefficients d_1, d_2, d_3, d_4, which occur in the factorial representation of the given number. Once these co-efficients are determined, the *tāna* or the permutation can be obtained by the standard procedure discussed earlier.

2.5. Historical Note

Among the available texts, *Saṅgītaratnākara* seems to be the earliest one to discuss the subject of *tāna-prastāra.* The famous mathematical text *Gaṇitakaumudī* (c. 1356) of Nārāyaṇa Paṇḍita has a chapter on *Aṅkapāśa.* Here, Nārāyaṇa gives the same rule for generating the *prastāra* of the set of all permutations of p digits (*Gaṇitakaumudī* XIII. 45-55). He also uses the same *khaṇḍa-meru* in order to discuss the *naṣṭa* and *uddiṣṭa* processes. As we have also remarked, the *naṣṭa* process is essentially based on a factorial representation of integers.

Though, the text does not mention it explicitly, the rule for *prastāra* given in *Saṅgītaratnākara* is also applicable for *tānas* with repetition of *svaras.* Nārāyaṇa discusses this case explicitly, along with the corresponding *naṣṭa* and *uddiṣṭa* processes (*Gaṇitakaumudī* XIII. 55-61).

In the European tradition, an algorithm for permutation generation is said to have been first given by C. F. Hindenburg in 1784 (D. E. Knuth: *The Art of Computer Programming Vol. IV, Fasc IV, History of Combinatorial Generation*, Addison Wesley, New York 2006, p. 61–2).

3. *Tāla-prastāra*

Apart from *Saṅgīta-samayasāra* of Pārśvadeva (c.1165), which has a brief discussion of *tāla-prastāra*, *Saṅgītaratnākara* is the oldest text which deals with *tāla-prastāra*. Its treatment of the subject is much more detailed and definitive than that found even in several later texts such as *Saṅgītopaniṣad-sāroddhāra* of Sudhākarakalaśa (c.1325), *Nartana-nirṇaya* of Paṇḍarīkaviṭṭhala (c.1560) and *Saṅgīta-darpaṇa* of Catura Dāmodara (c.1600).

Chapter V of *Saṅgītaratnākara* is the *Tālādhyāya* with 409 verses. Starting with the definition of *tāla*, the first 235 verses deal with *mārga-tālas* (classical rhythm patterns). Verses 236–311 discuss 120 *deśī-tālas* (regional rhythm patterns). At the end of this discussion, it is noted that there are indeed very many such *tālas* and it would not be possible to display all of them. This sets the stage for the *prastāra-prakaraṇa* which takes up the remaining nearly 100 verses of the *Tālādhyāya*.

3.1. *Tālāṅgas*

The *tālāṅgas*, or the basic rhythm components, considered by Śārṅgadeva are four, namely, *druta*, *laghu*, *guru* and *pluta*, which may be taken to be of duration 1, 2, 4 and 6 respectively, in *druta* units. Thus the *tāla-prastāra* considered in *Saṅgītaratnākara* is sometimes referred to as *caturaṅga-tāla-prastāra*.[6]

This process of *tāla-prastāra* is a generalization of the process of *mātra-vṛtta-prastāra* discussed in the text of *Chandas-śāstra*. There, only *laghu* and *guru* are considered and they are of one and two time units respectively. *Gaṇita-kaumudī* of Nārayaṇa Paṇḍita (c.1356) discusses a different kind of generalization of the *mātrā-vṛtta-prastāra* of *Chandas-śāstra*, by including units which are also of 3,4,..., *q-mātrās* in the *prastāra*. The *tāla-prastāra* of *Saṅgītaratnākara* cannot however be subsumed under Nārayaṇa's generalization.

3.2. *Pratyayas*

Pratyayas are simple techniques to comprehend the diverse features of the large number of *tālas* which are noticed in practice, without having to enumerate them

[6] In a series of books, the well known musician and musicologist Vidwan Akella Mallikarjuna Sarma, formerly Principal, Government College of Music, Hyderabad, has given a detailed exposition of the theory of *tāla-prastāra* as expounded in *Saṅgītaratnākara*. According to Vidwan Sarma, there are some later texts of *Saṅgīta-sāstra* (as yet unpublished), which include some more *tālāṅgas*, such as *anu-druta* etc., and different kinds of *laghus* (*khaṇḍa-jāti*, etc.) in their discussion of *tāla-prastāras*. It is important to bring these texts to light in order to understand further developments of the theory of *tāla-prastāra* due to the successors of Śārṅgadeva.

explicitly each time. Śārṅgadeva first lists the nineteen *pratyayas* considered in this *śāstra*.

अन्येऽपि सन्ति भूयांसस्तालास्ते लक्ष्यवर्त्मनि ॥
प्रसिद्धिविधुरत्वेन शास्त्रेऽस्मिन्न प्रदर्शिताः ।
तद्भेदप्रत्ययार्थं तु लघूपाया भवन्त्यमी॥
प्रस्तारसंख्ये नष्टं चोद्दिष्टं पातालकस्ततः।
द्रुतमेरुर्लघोर्मेरुर्गुरुमेरुः प्लुतस्य च ॥
मेरुः संयोगमेरुश्च खण्डप्रस्तारकस्ततः।
प्राचां चतुर्णां मेरुणां नष्टोद्दिष्टे पृथक्पृथक्॥
एकोनविंशतिरिति प्रत्ययास्तान् ब्रुवेऽधुना।

(सङ्गीतरत्नाकरः ५.३१२–३१६)

The *pratyayas* which are being discussed by Śārṅgadeva are the following:

- *Prastāra*: Systematic generation and enumeration of all possible *tāla*-forms with a given total number of time-units.
- *Sāṃkhyā*: Determination of the number of *tāla*-forms in a *prastāra*.
- *Naṣṭa*: Ranking of a given *tāla*-form in a *prastāra*.
- *Uddiṣṭa*: Finding the *tāla*-form corresponding to a given rank.
- *Pātāla*: Total number of *drutas*, *laghus*, *gurus* and *plutas* in a *prastāra*.
- *Druta-meru*, *Laghu-meru*, *Guru-meru* and *Pluta-meru*: Finding the number of *tāla*-forms with a given number of *drutas*, *laghus*, *gurus* and *plutas*, in a *prastāra*.
- *Sāṃyoga-meru*.
- *Khaṇḍa-prastāra*.
- *Uddiṣṭa* (and *naṣṭa*) ranking (and un-ranking) a particular *tāla*-form among those with a given number of *drutā*, *laghus*, *gurus* or *plutas*.

3.3. *Prastāra-vidhi* (Rules for the formation of the *Prastāra*)

न्यस्याल्पमाद्यान्महतोऽधस्ताच्छेषं यथोपरि।
प्रागूने वामसंस्थांस्तु संभवे महतो लिखेत्॥
अल्पानसंभवे तालपूर्त्यै भूयोऽप्ययं विधिः ।
सर्वद्रुतावधिः कार्यः प्रस्तारोऽयं लघौ गुरौ ॥
प्लुते व्यस्ते समस्ते च न तु व्यस्ते द्रुतेऽस्ति सः ।

(सङ्गीतरत्नाकरः ५.३१६–३१८)

Śārṅgadeva's procedure for the construction of *prastāra* is as follows:

- The last row of the *prastāra* has all *drutas* only.
- In the first row, place as many *plutas* as possible to the right, followed (from right to left), if possible, by a *guru* and a *druta* or a *guru* alone; or by a *laghu* and a *druta* or a *laghu* alone; or by a *druta* alone to the left.
- To go from any row of the *prastāra* to the next, identify the first non-*druta* element from the left. Place below that the element next to it with lower duration: Place *druta* below a *laghu*, *laghu* below a *guru* and *guru* below a *pluta*.
- Bring down the elements to the right as they are.
- Make up for the deficient units (if any) by adding to the left as many *plutas* as possible, followed similarly by *gurus*, *laghus* and *drutas* in that order from right to left.

Example: ***Sapta-druta-prastāra***

We shall use the symbols D, L, G, P to stand for *druta*, *laghu*, *guru* and *pluta* respectively. The traditional symbols for these are somewhat like o, |, S, Ś. In Table 6, we display the *sapta-druta-prastāra* or the enumeration of all rhythmic patterns of seven time units following the rule of enumeration of Śārṅgadeva. We also show the time duration of each *tālāṅga* on the right.

3.4. *Saṃkhyā*

Śārṅgadeva next discusses the *pratyaya* of *saṃkhyā*, which gives the number of rows in the *prastāra* associated with rhythms of a given time measure.

एकद्व्यङ्कौ क्रमान्न्यस्य युञ्जीतान्त्यं पुरातनैः।
द्वितीयतुर्यषष्ठाङ्कैरभावे तुर्यषष्ठयोः॥
तृतीयपञ्चमाङ्काभ्यां क्रमात् तं योगमग्रतः।
लिखेद् दक्षिणसंस्थैवमङ्कश्रेणी विधीयते ॥
सा चाङ्कैरिष्टतालस्थद्रुतसंख्यैः समाप्यते।
द्रुतो लघुः सार्धमात्रो गुरुः सार्धद्विमात्रिकः ॥
प्लुतः सार्धत्रिमात्रश्चेत्येकैकद्रुतवर्धितैः।
तालभेदाः क्रमादङ्कैः संख्यायन्ते स्थितैरिह॥
यदङ्कयोगादन्त्योऽङ्को लब्धस्तैरन्ततः क्रमात्।
भेदा द्रुतान्तलघ्वन्तगुर्वन्ताश्च प्लुतान्तकाः॥
संख्यायन्त इति प्रोक्ताः संख्या निःशङ्कसूरिणा।

(सङ्गीतरत्नाकरः ५.३१९-३२४)

Sl. No.	*Tāla* forms: In terms of *tālāṅgas*							In terms of *kāla-pramāṇas*						
1						D	P						1	6
2					D	L	G					1	2	4
3					L	D	G					2	1	4
4				D	D	D	G				1	1	1	4
5					D	G	L					1	4	2
6				D	L	L	L				1	2	2	2
7				L	D	L	L				2	1	2	2
8			D	D	D	L	L			1	1	1	2	2
9					G	D	L					4	1	2
10				L	L	D	L				2	2	1	2
11			D	D	L	D	L			1	1	2	1	2
12			D	L	D	D	L			1	2	1	1	2
13			L	D	D	D	L			2	1	1	1	2
14		D	D	D	D	D	L		1	1	1	1	1	2
15						P	D						6	1
16					L	G	D					2	4	1
17				D	D	G	D				1	1	4	1
18					G	L	D					4	2	1
19				L	L	L	D				2	2	2	1
20			D	D	L	L	D			1	1	2	2	1
21			D	L	D	L	D			1	2	1	2	1
22			L	D	D	L	D			2	1	1	2	1
23		D	D	D	D	L	D		1	1	1	1	2	1
24				D	G	D	D				1	4	1	1
25			D	L	L	D	D			1	2	2	1	1
26			L	D	L	D	D			2	1	2	1	1
27		D	D	D	L	D	D		1	1	1	2	1	1
28				G	D	D	D				4	1	1	1
29			L	L	D	D	D			2	2	1	1	1
30		D	D	L	D	D	D		1	1	2	1	1	1
31		D	L	D	D	D	D		1	2	1	1	1	1
32		L	D	D	D	D	D		2	1	1	1	1	1
33	D	D	D	D	D	D	D	1	1	1	1	1	1	1

TABLE 6. Seven-*druta-prastāra*.

The sequence of *saṃkhyāṅkas* S_n, where S_n is the number of *tāla* forms in the *n-druta-prastāra*, is obtained as follows: The first two elements of the sequence are $S_1 = 1, S_2 = 2$. The *n*-th element of the sequence (S_n) is obtained by

adding the previous (S_{n-1}), second-previous (S_{n-2}), fourth-previous (S_{n-4}) and the sixth-previous (S_{n-6}) elements. That is, the *saṃkhyāṅkas* satisfy the recurrence relation:

$$S_n = S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6}$$

In case the fourth or sixth-previous elements do not exist, the third and the fifth-previous (if they exist) are to be taken as their representatives (*pratinidhi*). As we shall see, this happens only in two cases, namely S_4, S_6.

Śārṅgadeva also makes the very interesting observation that among all the *tāla*-forms which appear in the *n-druta-prastāra*, S_{n-1} end in a *druta*, S_{n-2} in a *laghu*, S_{n-4} in a *guru* and S_{n-6} end in a *pluta*, and the total number of *tāla*-forms S_n in the *n-druta-prastāra* is just the sum of these four numbers. We shall subsequently discuss the significance of this observation.

n	1	2	3	4	5	6	7	8	9	10
S_n	1	2	3	6	10	19	33	60	106	191

TABLE 7. Sequence of *saṃkhyāṅkās*.

In Table 7, we give the sequence of *saṃkhyāṅkas* S_n. Here, $S_1 = 1, S_2 = 2$, $S_3 = 2+1 = 3$, $S_4 = 3+2+1 = 6$, $S_5 = 6+3+1 = 10$, $S_6 = 10+6+2+1 = 19$, $S_7 = 19+10+3+1 = 33$, $S_8 = 33+19+6+2 = 60$, and so on. The number 1 which appears in the sum for obtaining S_4 and S_6, is the third-previous and the fifth-previous *saṃkhyāṅka* respectively, and is used here as the representative (*pratinidhi*) of the fourth-previous and the sixth-previous *saṃkhyāṅkas*, as they are absent. It is sometimes suggested that we can get rid of this prescription to use the representatives, by introducing an entry $S_0 = 1$, before the sequence of *saṃkhyāṅkas*.[7] The recursion relation $S_n = S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6}$ for the *saṃkhyāṅkas* is a generalization of the recursion relation

$$s_n = s_{n-1} + s_{n-2}$$

that defines the *saṃkhyāṅkas* s_n in a *mātrā-vṛtta-prastāra*. The latter relation was noted by Virahāṅka (c. 650) and gives rise to the so called Fibonacci sequence or, more appropriately, the Virahāṅka-Fibonacci sequence of numbers.[8] Indeed the mathematical theory of *tāla-prastāra* turns out to be a very interesting

[7] See for instance, the works of Akella Mallikarjuna Sarma cited in the bibliography.

[8] In *Gaṇitakaumudī* (c.1356), Nārāyaṇa Paṇḍita discusses a different kind of generalization of the Virahāṅka-Fibonacci sequence by considering the recurrence relation $S_n = S_{n-1} + S_{n-2} + \dots + S_{n-q}$ to generate what he calls as the *sāmāsikī paṅkti* of order q.

1				G
2			L	L
3		D	D	L
4		D	L	D
5		L	D	D
6	D	D	D	D

TABLE 8. Four-*druta-prastāra.*

generalization of the mathematical theory of the *mātrā-vṛtta-prastāra* as we shall see below.

3.5. An inductive construction of the *Tāla-prastāra* and the *Saṃkhyā* rule

The construction process of *tāla-prastāra* is indeed a generalization of the construction of *mātrā-vṛtta-prastāra* as described in texts of Sanskrit prosody. Whilst *laghu* (l) and *guru* (g) are units of value 1 and 2 respectively in prosody, in the theory of musical rhythms there are four units *druta* (D), *laghu* (L), *guru* (G) and *pluta* (P) which have values 1, 2, 4 and 6 respectively.

In prosody, a row of a *prastāra*, of a given value n, consists of a sequence of *laghus* (l) and *gurus* (g) such that the total value of the row is n, where the total value is by definition the sum of all the values of the symbols in the row. The *prastāra* is actually an enumeration of all sequences of l and g by a certain rule, such that each row has value n, and all possible rows of value n are included under the rule of enumeration. The construction of the *prastāra* is achieved by an inductive process.

A similar process is at work in the case of *tāla-prastāra.* The *prastāra* of value 1 is just D. So the corresponding *saṃkhyā* (the number of *tāla*-patterns corresponding to a given value) is $S_1 = 1$. The *prastāra* of value 2, consists of two rows, L and DD and correspondingly $S_2 = 2$. The *prastāra* of value 3, consists of three rows, DL, LD and DDD and correspondingly $S_3 = 3$. The *prastāra* of value four has six rows as shown in Table 8, and correspondingly $S_4 = 6$.

The *prastāra* of value five has ten rows as shown in Table 9, and correspondingly $S_5 = 10$. The *prastāra* of value six has nineteen rows as shown in Table 10. Thus $S_6 = 19$. In each of the above cases, we have an enumeration of all rows of value $1, 2, 3, 4, 5$ and 6.

It is now clear that the *tāla-prastāra* of value n can be constructed in the following manner. We first take the *prastāra* consisting of rows of value $n-6$ and augment it by adding a *pluta* (P) at the end. Below that, place the *prastāra* of value $n-4$ and augment it by placing a *guru* (G) at the end. Below that, place the

1				D	G
2			D	L	L
3			L	D	L
4		D	D	D	L
5				G	D
6			L	L	D
7		D	D	L	D
8		D	L	D	D
9		L	D	D	D
10	D	D	D	D	D

TABLE 9. Five-*druta-prastāra.*

1						P
2					L	G
3				D	D	G
4					G	L
5				L	L	L
6			D	D	L	L
7			D	L	D	L
8			L	D	D	L
9		D	D	D	D	L
10				D	G	D
11			D	L	L	D
12			L	D	L	D
13		D	D	D	L	D
14				G	D	D
15			L	L	D	D
16		D	D	L	D	D
17		D	L	D	D	D
18		L	D	D	D	D
19	D	D	D	D	D	D

TABLE 10. Six-*druta-prastāra.*

prastāra of value $n-2$ and augment it by placing a *laghu* (L) at the end. Below that, finally, place the *prastāra* of value $n-1$ and augment it by placing a *druta* (D) at the end. We can see the above process clearly at work by noting that the *prastāra* of value 7, given in Table 6, is constructed in the manner described above from *prastāras* of values 1, 3, 5 and 6.

3.6. The generating function for *Saṃkhyāṅkas* S_n

We first explain how, in the case of *mātrā-vṛtta-prastāra* in prosody, the sequence of *saṃkhyās* (the number of moric metres of a given value) may be obtained. For example, let us consider the row lgg which occurs in the *prastāra* of rows of value 5. We associate a term in the expansion $(x+x^2)^3$ namely the term x^5 which is obtained by taking x from the first factor and x^2 from the second and third factors and multiplying the monomials so chosen. To consider another example, to the row $glll$, we associate a term in the expansion of $(x+x^2)^4$ namely the term x^5 which is obtained similarly by taking x^2 from the first factor and x from the second, third and fourth factors. From such examples it is clear that if s_n is the number of rows in the *mātrā-vṛtta-prastāra* of value n, then the generating function of s_n is given by

$$\begin{aligned}\sum s_n x^n &= 1+(x+x^2)+(x+x^2)^2+\ldots(x+x^2)^r+\ldots \\ &= [1-(x+x^2)]^{-1}\end{aligned}$$

It turns out that the number s_n is the so called n-th Fibonacci number, or what may more appropriately be called the n-th Virahāṅka-Fibonacci number.

It follows in a similar manner that the generating function for the *saṃkhyāṅkas* S_n (the number of rows in the *tāla-prastāra* of value n) is given by

$$\begin{aligned}\sum S_n x^n &= 1+(x+x^2+x^4+x^6)+(x+x^2+x^4+x^6)^2+\ldots \\ &\qquad +(x+x^2+x^4+x^6)^r+\ldots \\ &= [1-(x+x^2+x^4+x^6)]^{-1} \qquad (5)\end{aligned}$$

Therefore, we have

$$[1-(x+x^2+x^4+x^6)]\sum S_n x^n = 1.$$

Equating coefficients of the low powers of x, we get $S_1 = S_0 = 1,\ S_2 = S_1 + S_0 = 2,\ S_3 = S_2 + S_1 = 3,\ S_4 = S_3 + S_2 + S_0 = 6$, etc. Further, equating coefficients of x^n for $n \geq 6$, we get

$$S_n = S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6} \text{ for } n \geq 6. \qquad (6)$$

We may refer to the *sāṃkhyāṅkas* S_n as the Śārṅgadeva numbers. These numbers satisfy a more complex recurrence relation (6) as compared with the simple recurrence relation

$$s_n = s_{n-1} + s_{n-2} \qquad (7)$$

that characterizes the Virahāṅka-Fibonacci numbers.

3.7. The decomposition of an integer in terms of Śārṅgadeva numbers S_n

In this section we shall exhibit a procedure to decompose an integer $m \geq 1$, in terms of the Śārṅgadeva numbers S_n using induction. The motivation for this comes from the following result for the Fibonacci numbers which was demonstrated in the mathematical theory of *mātrā-vṛtta-prastāra.*[9]

Proposition: Any positive integer is either a Fibonacci number or can be expressed uniquely as a sum of non-consecutive Fibonacci numbers.

In what follows, we shall prove an analogue of this result for Śārṅgadeva numbers. If $m = 1$, we write $1 = S_1$. Then we write $2 = S_2,\ 3 = S_3,\ 4 = S_1 + S_3$. Now, for any general m, choose the largest integer q such that $S_q \leq m$. By induction, following the same procedure, we can decompose $m - S_q$, that is write it in the form

$$m - S_q = S_{i_1} + S_{i_2} + \ldots + S_{i_p} \qquad \text{with } i_1 \leq i_2 \leq \ldots \leq i_p \leq q$$

Then,

$$m = S_{i_1} + S_{i_2} + \ldots + S_{i_p} + S_q \qquad \text{with } i_1 \leq i_2 \leq \ldots \leq i_p \leq q \tag{8}$$

is a decomposition of m which we shall refer to as the canonical decomposition. Here, for any m, with $1 \leq k \leq p$, i_k is the largest integer such that $S_{i_k} \leq (m - (S_{i_{k+1}} + S_{i_{k+2}} + \ldots + S_{i_p} + S_q))$. We may also note that if m is itself a Śārṅgadeva number S_q, then $m = S_q$ is the canonical decomposition of m. We shall now show that the canonical decomposition has certain properties. For this we need the following lemma.

Lemma 1: The Śārṅgadeva numbers S_n satisfy the following relations:

$$(i) \qquad 2S_n \geq S_{n+1}, \text{ and equality holds only if } n = 1, 3 \tag{9}$$

$$(ii) \qquad S_n + S_{n+1} + S_{n+2} \geq S_{n+3}, \text{ and equality holds only if } n = 1 \tag{10}$$

$$(iii) \qquad S_n + S_{n+1} + S_{n+3} + S_{n+4} \geq S_{n+5}, \text{ and equality holds only if } n = 1 \tag{11}$$

Proof: These relations follow directly from the recurrence relation (6). We have

$$S_n = S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6}$$

[9] The mathematical theory of *mātrā-vṛtta-prastāra* is outlined in R. Sridharan, *Pratyayas for Mātrāvṛttas* and Fibonacci Numbers, Mathematics Teacher, 42, 2006, 120-137.

We also note that for $n \geq 1$, $S_n \geq S_{n-1}$. Hence

$$\begin{aligned} 2S_n &= 2S_{n-1} + 2S_{n-2} + 2S_{n-4} + 2S_{n-6} \\ &\geq S_n + S_{n-1} + S_{n-3} + S_{n-5} \qquad \text{(by induction)} \\ &= S_{n+1}. \end{aligned}$$

This proves (i). Now we prove (ii) by induction. We note that property (ii) holds if $n = 2$. Now, we use the recurrence relation (6) to write

$$\begin{aligned} S_n + S_{n+1} + S_{n+2} &= (S_{n-1} + S_n + S_{n+1}) + (S_{n-2} + S_{n-1} + S_n) \\ &\quad +(S_{n-4} + S_{n-3} + S_{n-2}) + (S_{n-6} + S_{n-5} + S_{n-4}) \\ &\geq S_{n+2} + S_{n+1} + S_{n-1} + S_{n-3} \qquad \text{(by induction)} \\ &= S_{n+3} \end{aligned}$$

This proves (ii). Property (iii) follows directly from the recurrence relation (6), as we have

$$S_n + S_{n+1} + S_{n+3} + S_{n+4} \geq S_{n-1} + S_{n+1} + S_{n+3} + S_{n+4} = S_{n+5}$$

This completes the proof of the lemma.

Now consider some number $m < S_n$ and let

$$m = S_{i_1} + S_{i_2} + \ldots + S_{i_t} \qquad \text{with } i_1 < i_2 < \ldots < i_t \tag{12}$$

be the canonical decomposition of the number, where we have taken note of the fact (which follows from (9) and the construction of the canonical decomposition) that each Śārṅgadeva number S_i appears at the most once in the canonical decomposition of a number. We can therefore write the above decomposition in the form

$$m = \sum \alpha_i S_i, \qquad \text{with } \alpha_i = 0 \text{ or } 1, \qquad \text{for } 1 \leq i < n, \qquad \text{and } \alpha_n = 0 \tag{13}$$

Hence, to each such decomposition (12), we assign a binary sequence $(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}, 0)$ of zeroes and ones of length n in the following manner. With the number m, which has the canonical decomposition (12), we associate the sequence of length n with zeroes at all places except at the positions $i_1 < i_2 < \ldots < i_t$ where it has ones. From the above lemma, we can easily see that such a binary sequence of length n (ending with 0), associated with any number $m < S_n$, has the following properties:

(a) A combination of the type (1, 1, 1), that is three 1s in a succession, cannot occur as a part of the above sequence.

(b) A combination of the type (1, 1, 0, 1, 1) cannot occur as a part of the above sequence.

(c) A combination of the type (1, 0, 1, 0, 1, 1) cannot occur as a part of the above sequence [This follows from the recurrence relation (6)].

(d) The first two entries cannot both be 1s in the above sequence [This follows from the fact that the first three Śārṅgadeva numbers are related by $S_3 = S_1 + S_2$].

(e) The sequence cannot begin with the combination (1,0,1,1,0) [This follows from the fact that $S_5 = S_1 + S_3 + S_4$].

We now proceed to show in Proposition 3 that the above conditions are sufficient to characterise the binary sequences that are associated with the canonical decomposition of a number. We first prove a lemma that is needed for the proof of this proposition.

Lemma 2: Let

$$m = S_{i_1} + S_{i_2} + \ldots + S_{i_t} \qquad \text{with } i_1 < i_2 < \ldots < i_t$$

be any decomposition of a number into Śārṅgadeva numbers and assume that the corresponding binary sequence of 0's and 1's satisfies the conditions (a)-(e). Then

$$S_{i_1} + S_{i_2} + \ldots + S_{i_t} < S_{i_t+1} \tag{14}$$

Proof: We shall prove (14) by induction. We note that (14) is equivalent to

$$S_{i_1} + S_{i_2} + \ldots + S_{i_{(t-1)}} < S_{i_t+1} - S_{i_t} \tag{15}$$

(which we know is $S_{i_t-1} + S_{i_t-3} + S_{i_t-5}$).

If $i_{t-1} < i_t - 1$, then by induction,

$$S_{i_1} + S_{i_2} + \ldots + S_{i_{(t-1)}} < S_{i_{(t-1)}+1} \leq S_{i_t-1}$$

and we are through with (15) and hence with (14). Suppose, $i_{t-1} = i_t - 1$. Then, we can cancel both $S_{i_{(t-1)}} = S_{i_t-1}$ from both sides of (15) and we are left to show that

$$S_{i_1} + S_{i_2} + \ldots + S_{i_{(t-2)}} < S_{i_t-3} + S_{i_t-5}. \tag{16}$$

If $i_{t-2} < i_t - 3$, then we are through by induction. Otherwise, since $i_{t-2} < i_{t-1} < i_t$ and $i_{t-1} = i_t - 1$, i_{t-2} is either $i_t - 2$ or $i_t - 3$. If $i_{t-2} = i_t - 2$, then since $i_{t-1} = i_t - 1$, the combination (1, 1, 1) appears in the sequence, which is a contradiction. Thus, the only possibility is $i_{t-2} = i_t - 3$, in which case, we can

cancel $S_{i_{(t-2)}}$ from both sides of (16) and are left to prove

$$S_{i_1} + S_{i_2} + \ldots + S_{i_{(t-3)}} < S_{i_t - 5} \tag{17}$$

Now, $i_{t-3} < i_{t-2} = i_t - 3$. Suppose, $i_{t-3} < i_t - 5$, then we are through by induction as before. Otherwise, we have $i_{t-3} \geq i_t - 5$; then since $i_{t-3} < i_{t-2} = i_t - 3$, we have that i_{t-3} is either $i_t - 4$ or $i_t - 5$. If, $i_{t-3} = i_t - 4$, then the binary sequence will have the combination $(1, 1, 0, 1, 1)$ which has been excluded by assumption. On the other hand, if $i_{t-3} = i_t - 5$, then the binary sequence will have the combination $(1, 0, 1, 0, 1, 1)$ which has also been excluded by assumption. Hence the lemma is proved.

Proposition 3: Let

$$m = S_{i_1} + S_{i_2} + \ldots + S_{i_t} \quad \text{with } i_1 < i_2 < \ldots < i_t \tag{18}$$

be any decomposition of a number and assume that the corresponding binary sequence of 0's and 1's satisfies the conditions (a)-(e). Then the above decomposition is the canonical decomposition of m. Equivalently, let

$$\begin{aligned} m &= S_{i_1} + S_{i_2} + \ldots + S_{i_t} \quad \text{with } i_1 < i_2 < \ldots < i_t \\ &= S_{j_1} + S_{j_2} + \ldots + S_{j_r} \quad \text{with } j_1 < j_2 < \ldots < j_r \end{aligned} \tag{19}$$

be any two decompositions of the number m, such that the corresponding binary sequences satisfy the conditions (a)-(e). Then the two decompositions in (19) are the same.

Proof: If $i_t = j_r$, then we can cancel $S_{i_t} = S_{j_r}$ from both the decompositions and use induction to prove the proposition. Hence we can assume without loss of generality that $i_t < j_r$. Now, from the above lemma it follows that

$S_{i_1} + S_{i_2} + \ldots + S_{i_t} < S_{i_t+1}$

But, since $i_t < j_r$ it follows that $S_{i_t+1} \leq S_{j_r}$. Hence, we get

$S_{i_1} + S_{i_2} + \ldots + S_{i_t} < S_{j_r}$ which contradicts the assumption (19). Hence the proposition is proved.

Corollary: The above propositions establish a bijection between the set of non-negative integers $0, 1, 2, \ldots, S_n - 1$, which label the rows in a *n-druta-prastāra*, and binary sequences of length n (ending with 0), which satisfy the conditions (a)-(e) stated above.

3.8. The mathematical basis of the *naṣṭa* and *uddiṣṭa* process

We enumerate the rows of a *prastāra* of a total value n starting from the last row, numbering the last row as 0, the penultimate row as 1, and so on. If the rank of a row in this numbering is t, then the number of the row in the usual enumeration (starting from the top and calling the top row as row number 1) is $S_n - t$. The

process of *naṣṭa* consists in writing down explicitly the row (without the aid of the *prastāra*), if one is given the row number m in the enumeration we have just described. If $m = 0$, then the row consists of n Ds. If $m < S_{n-1}$ then the last symbol of the row is D because the last S_{n-1} rows of the n-*druta prastāra* end with a D, by very definition. If $S_{n-1} \leq m < S_{n-1} + S_{n-2}$, then the m-th row ends with an L. If $S_{n-1} + S_{n-2} \leq m < S_{n-1} + S_{n-2} + S_{n-4}$, then the row ends with a G. If $S_{n-1} + S_{n-2} + S_{n-4} \leq m < S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6}$, then the row ends with a P.

This leads one to the expectation that the canonical decomposition of $m(1 \leq m < S_n)$, determines explicitly the sequence of symbols that occur in the m-th row. To show this, we proceed inductively as follows.

(i) If the last symbol is D, that is $m < S_{n-1}$, then the penultimate symbol of the m-th row is the last symbol of the m-th row in the *prastāra* of value $n - 1$, where the last row has $n - 1$ *drutas*.

(ii) If the m-th row ends with an L, that is $S_{n-1} \leq m < S_{n-1} + S_{n-2}$, then the penultimate symbol of the m-th row is the last symbol of the $(m - S_{n-1})$-th row of the *prastāra* of value $n - 2$.

(iii) If the m-th row ends with a G, that is $S_{n-1} + S_{n-2} \leq m < S_{n-1} + S_{n-2} + S_{n-4}$, then the penultimate symbol of the m-th row is the last symbol of the $(m - (S_{n-1} + S_{n-2}))$-th row of the *prastāra* of value $n - 4$.

(iv) If the m-th row ends with a P, that is $S_{n-1} + S_{n-2} + S_{n-4} \leq m < S_{n-1} + S_{n-2} + S_{n-4} + S_{n-6}$, then the penultimate symbol of the m-th row is the last symbol of the $(m - (S_{n-1} + S_{n-2} + S_{n-4}))$-th row of the *prastāra* of value $n - 6$.

The next or the second penultimate symbol is then determined in the same way. And so on. This essentially establishes the following relation between the canonical decomposition of the row number m and the sequence of symbols that occur in the row.

Proposition 4: Given any number $0 \leq m < S_n$ let the corresponding canonical decomposition be obtained and re-expressed as a binary sequence of $0'$s and $1'$s in the manner described earlier. Then, the symbols in the m-th row (enumerated from below starting from the last row numbered as 0) in the *prastāra* of value n, are as follows:

(i) If the sequence ends with $(0, 0)$, then the last symbol is a D and the next symbol is the last symbol of the row whose number is given by the truncated binary sequence with the last 0 removed, in the *prastāra* of value $n - 1$.

(ii) If the sequence ends with $(0, 1, 0)$, then the last symbol is an L and the next symbol is the last symbol of the row whose number is given by the truncated

binary sequence with the end portion $(1, 0)$ removed, in the *prastāra* of value $n-2$.

(iii) If the sequence ends with $(0, 0, 1, 1, 0)$, then the last symbol is a G and the next symbol is the last symbol of the row whose number is given by the truncated binary sequence with the end portion $(0, 1, 1, 0)$ removed, in the *prastāra* of value $n-4$.

(iv) If the sequence ends with $(0, 0, 1, 0, 1, 1, 0)$, then the last symbol is a P and the next symbol is the last symbol of the row whose number is given by the truncated binary sequence with the end portion $(0, 1, 0, 1, 1, 0)$ removed, in the *prastāra* of value $n-6$.

We now look at the converse process of *uddiṣṭa*, where the objective is to determine the row-number (in the enumeration given above) given the sequence of symbols that occur in some row of the *prastāra* of value n. Suppose the row-number is m and we want to determine m. Suppose, the last symbol is D, then $m < S_{n-1}$ and therefore S_{n-1} does not occur in the canonical decomposition of m. If we delete the last D, we obtain a row of value $n-1$, whose row-number t in the *prastāra* (of value $n-1$), can be determined by induction and we have $m = t$. Suppose the last symbol is L, then S_{n-1} occurs in the decomposition for m, and after deleting L, we get row of value $n-2$, whose row-number t in the *prastāra* (of value $n-2$) can be determined by induction on n. We then set $m = S_{n-1} + t$. If the last symbol of the m-th row is G, then S_{n-1} and S_{n-2} occur in the decomposition of m, and if we delete this G, we obtain a row of value $n-4$, whose row-number t in the *prastāra* (of value $n-4$) can be determined by induction and we set $m = S_{n-1} + S_{n-2} + t$. If the last symbol of the m-th row is P, then S_{n-1}, S_{n-2} and S_{n-4} occur in the decomposition of m, and if we delete this P, we obtain a row of value $n-6$, whose row-number t in the *prastāra* (of value $n-6$) can be determined by induction and then we set $m = S_{n-1} + S_{n-2} + S_{n-4} + t$. This is the process of *uddiṣṭa*.

We have outlined the theory of the canonical representation of a number as a sum of Śārṅgadeva numbers (*saṃkhyāṅkas* S_n), which constitutes the mathematical basis for the *naṣṭa* and *uddiṣṭa* processes discussed by Śārṅgadeva. We shall now outline the processes of *naṣṭa* and *uddiṣṭa* as described by Śārṅgadeva himself. His rule of enumeration of *tāla-prastāra* also has an elegant mathematical description if viewed in terms of the canonical decompositions associated with two successive row-numbers. We shall discuss this in Appendix A.

3.9. *Naṣṭa* process as discussed by Śārṅgadeva

अत्रैतावतिथो भेदः किंरूप इति पृच्छति ॥
यत्र तन्नष्टमाख्यातं तस्योत्तरमिहोच्यते।

भेदानां यावतां मध्ये नष्टप्रश्नः कृतो भवेत्।
तावत्संख्याङ्कपर्यन्तां लिखेत् संख्याङ्कसंततिम्।
अन्त्याङ्के तत्र नष्टाङ्कं पातयेदथ शेषतः ॥
पातयेत् पूर्वपूर्वाङ्कं तत्रत्त्वपतितो द्रुतः।
पूर्वश्चेत्पतितो न स्याल्लघुस्तु पतिताद्भवेत् ॥
उत्तरेणाकृतार्थेन सहितात् तदसंभवे ।
अकृतार्थेन पूर्वेण सान्तरे पतिते पृथक्॥
लघुर्निरन्तरे त्वस्मिँल्लघुरेव गुरुर्भवेत् ।
गुरुहेतोस्तृतीये तु पतिते गः प्लुती भवेत्॥
अङ्काभावे द्रुता ग्राह्यास्तालपूरणहेतवः॥
इति नष्टस्य विज्ञेयमुत्तरं रूपनिर्णयात्॥

(सङ्गीतरत्नाकरः ५.३२४–३३०)

The *naṣṭa* process is outlined by Śārṅgadeva in the above verses, which may be paraphrased as follows: In order to find the *tāla*-form which is associated with the row-number r (*naṣṭāṅka*) of a *n-druta-prastāra*, write the sequence of *saṃkhyāṅkas* $S_1, S_2, \ldots S_n$. Subtract the rank-number from S_n. From the result $(S_n - r)$, subtract S_{n-1} if possible. If S_{n-1} can be subtracted, it is called reducible (*patita*), and then subtract S_{n-2} (if possible) from the result $(S_n - r - S_{n-1})$ of this subtraction. If S_{n-1} cannot be subtracted from $S_n - r$, it is called irreducible (*apatita*) and then subtract S_{n-2} (if possible) from the number $S_n - r$. And so on.

Mark all the *saṃkhyāṅkas* $1, 2 \ldots S_{n-1}$ as either *patita* or *apatita*. S_n, being larger than r, is always *apatita*. Scanning the *saṃkhyāṅka* sequence right to left from S_n, obtain the successive *tālāṅgas* of the *tāla*-form from right to left, according to the following procedure.

- We start with S_n which is *apatita*. Then there are the following possibilities:
- If the following *saṃkhyāṅka* S_{n-1} is also *apatita*, then write a D corresponding to S_n and proceed with scanning from S_{n-1}.
- If S_{n-1} is *patita* and S_{n-2} is *apatita*, then write an L corresponding to the two *saṃkhyāṅkas* S_n and S_{n-1} taken together, and proceed to scan from S_{n-2}.
- If the *saṃkhyāṅkas* S_{n-1}, S_{n-2} are *patita* and S_{n-3} and S_{n-4} are *apatita*, then write a G corresponding to the four *saṃkhyāṅkas* S_n, S_{n-1}, S_{n-2} and S_{n-3} taken together, and start scanning from S_{n-4}.
- If the *saṃkhyāṅkas* $S_{n-1}, S_{n-2}, S_{n-4}$ are all *patita* and S_{n-3}, S_{n-5} and S_{n-6} are *apatita*, then mark a P corresponding to the six *saṃkhyāṅkas* $S_n, S_{n-1}, S_{n-2}, S_{n-3}, S_{n-4}$ and S_{n-5} taken together, and start scanning from S_{n-6}.

- Given the way the *saṃkhyāṅkas* are generated, the above four cases exhaust all the possibilities.
- Proceeding thus, mark all the *tālāṅgas* from right to left.

In the above process of *naṣṭa* for finding the *tāla*-form associated with row r in a *n-druta-prastāra*, Śārṅgadeva essentially gives the procedure for constructing what we have described as the canonical decomposition of the number $S_n - r$. If we mark the *patita saṃkhyāṅkas* by 1 and the *apatita* by 0, then following the procedure of Śārṅgadeva, we obtain the binary sequence that is associated with the number $S_n - r$ via its canonical decomposition. Then Śārṅgadeva's prescription for identifying the *tālāṅgas* that occur in the *tāla*-form appearing in the r-th row is essentially the same as what we have stated as Proposition 4 above (except for the fact that we have used an enumeration of the rows where the rows are numbered from 0 from the below, while Śārṅgadeva uses the usual enumeration of rows starting with row number 1 at the top of the *prastāra*).

The signatures of various *tālāṅgas* as given by Śārṅgadeva are given in the Table 11. Here, "0" stands for *apatita* and "1" for *patita*. Further, "(0)" (zero enclosed by brackets) indicates that the previous *saṃkhyāṅka* (if any) has to be an *apatita* for the identification to take effect.

	S_{n-6}	S_{n-5}	S_{n-4}	S_{n-3}	S_{n-2}	S_{n-1}	S_n
D						(0)	0
L					(0)	1	0
G			(0)	0	1	1	0
P	(0)	0	1	0	1	1	0

TABLE 11. Signatures of various *tālāṅgas*.

Examples

The 8^{th} *tāla-form in the 7-druta-prastāra*

$33 - 8 = 25,\ 25 - 19 = 6,\ 6 - 6 = 0$

Thus the *patita* and *apatita saṃkhyāṅkas* are as follows

0/1	0	0	0	1	0	1	0
S_n	1	2	3	6	10	19	33

- Starting from 33, since 19 is *patita* and 10 is *apatita*, we get a *laghu* at the right extreme.
- Starting from 10, since 6 is *patita* and 3 is *apatita*, we get another *laghu* to the left of the first.

- Starting from 3, since 2 is *apatita*, we get a *druta* to the left.
- Starting from 2, since 1 is *apatita* we get one more *druta* to the left.
- Since 1 is *apatita*, we get one more *druta*,

Thus the *tāla*-form is $DDDLL$.

The 28^{th} *tāla-form in the 7-druta-prastāra*

33, 19, 10 and 6 are a*patita* and thus we get DDD from the right. Starting from 6, 3 and 2 are *patita* and 1 is *apatita*. They give a G, thus giving the *tāla*-form $GDDD$.

0/1	0	1	1	0	0	0	0
S_n	1	2	3	6	10	19	33

3.10. *Uddiṣṭa* process as discussed by Śārṅgadeva

ईदृग्रूपोऽत्र कतिधा भेदः प्रश्न इतीदृशः ।
उद्दिष्टं तत्र संख्याङ्कसंततिं नष्टवल्लिखेत् ॥
यैरङ्कैः पतितैर्नष्टे लभ्यन्ते ये द्रुतादयः ।
तानेवाङ्कॉंल्लभन्ते ते भेदमुद्दिष्टमाश्रिताः ॥
यद्वा षट्प्लुतहेत्वङ्कमध्येऽन्त्यात् प्राचि सप्तमे ॥
तदभावे तु षष्ठाङ्के पातितेऽन्त्याङ्कमध्यतः ॥
यः शेषः स प्लुताल्लभ्यो लब्धहीनान्त्यशेषतः ॥
ज्ञानं पूरणसंख्याया उद्दिष्टोत्तरमिष्यते ॥

(सङ्गीतरत्नाकरः ५.३३१-३३४)

Given a *tāla*-form associated with an *n-druta-prastāra*, to find its rank-number the *uddiṣṭa* process is employed. In the above verses, Śārṅgadevā essentially states that since *uddiṣṭa* is the reverse of the *naṣṭa* process, we write the sequence of *saṃkhyāṅkas* and identify the *patitas* among them which give rise to the given sequence of *tālāṅgas*. This effectively leads to the following process, which has also been discussed earlier in the general mathematical context.

We write the *saṃkhyāṅkas* $1, 2, \ldots, S_n$ above the *tālāṅgas* from the left. We write one *saṃkhyāṅka* above a D, two above an L, four above a G and six above a P. Since the *saṃkhyāṅka* associated with a D is *apatita*, it is not marked. Since the first *saṃkhyāṅka* associated with an L is *patita* and the second is *apatita*, only the first is marked. Since only the second and third of the *saṃkhyāṅkas* associated with a G are *patita*, they are marked. Since, only the second, fourth and fifth *saṃkhyāṅkas* associated with a P are *patita*, they are marked. The sum of

the *patita* or marked *saṃkhyāṅkas* is subtracted from S_n to get the rank-number of the *tāla*-form.

Examples

To find the rank of LDLL in 7-druta-prastāra

<table>
<tr><td>1</td><td>2</td><td>3</td><td>6</td><td>10</td><td>19</td><td>33</td></tr>
<tr><td colspan="2">L</td><td>D</td><td colspan="2">L</td><td colspan="2">L</td></tr>
</table>

Total of the *patita saṃkhyāṅkas*: $19 + 6 + 1 = 26$. Rank: $33 - 26 = 7$.

To find the rank of GDL in 7-druta-prastāra

<table>
<tr><td>1</td><td>2</td><td>3</td><td>6</td><td>10</td><td>19</td><td>33</td></tr>
<tr><td colspan="3">G</td><td colspan="2">D</td><td colspan="2">L</td></tr>
</table>

Total of the *patita saṃkhyāṅkas*: $19 + 3 + 2 = 24$. Rank: $33 - 24 = 9$.

To find the rank of PD in 7-druta-prastāra

<table>
<tr><td>1</td><td>2</td><td>3</td><td>6</td><td>10</td><td>19</td><td>33</td></tr>
<tr><td colspan="6">P</td><td>D</td></tr>
</table>

Total of the *patita saṃkhyāṅkas*: $10 + 6 + 2 = 18$. Rank: $33 - 18 = 15$

To find the rank of DDLLD in 7-druta-prastāra

<table>
<tr><td>1</td><td>2</td><td>3</td><td>6</td><td>10</td><td>19</td><td>33</td></tr>
<tr><td>D</td><td>D</td><td colspan="2">L</td><td colspan="2">L</td><td>D</td></tr>
</table>

Total of the *patita saṃkhyāṅkas*: $10 + 3 = 13$. Rank: $33 - 13 = 20$.

3.11. *Pātāla*

आदौ रूपमथैकैकमङ्कसंख्यांकसंततेः।
क्रमादधोऽधो विन्यस्येदन्त्यादींश्चतुरस्तथा॥
स्वपङ्क्तिस्थांलिखेदङ्कानग्रसंख्याङ्कवद्द्रुतान्॥
किं तु प्रतिनिधिर्नात्र विद्यते तुर्यषष्ठयोः॥

इष्टतालद्रुतमितेष्वङ्केषु लिखितेष्विति।
क्रमादन्त्योपान्त्यतुर्यषष्ठैर्यत्र द्रुतादयः ॥
मीयन्ते सर्वभेदस्थाः पातालः सोऽभिधीयते।

(सङ्गीतरत्नाकरः ५.३३५-३३८)

The *pātāla* sequence $T_1, T_2, \ldots, T_n, \ldots$ is defined inductively as follows: First write 1 ($S_0 = 1$) followed by the sequence of *saṃkhyāṅkas* $S_1, S_2, \ldots, S_n, \ldots$. Below these the *Pātāla* sequence is to be set down. The n-th *pātāla* number (T_n) is obtained by adding the previous *saṃkhyāṅka* (S_{n-1}) to the previous, second-previous, fourth-previous and the sixth-previous *pātāla* numbers whenever they exist. That is,

$$T_n = S_{n-1} + T_{n-1} + T_{n-2} + T_{n-4} + T_{n-6} \qquad (20)$$

In the case of *pātāla* sequence, there is no consideration of any representative in case the fourth or sixth previous elements do not exist (which was done in the case of *saṃkhyaṅka* sequence).[10]

The *pātāla* numbers T_n, T_{n-1}, T_{n-3} and T_{n-5} give the total number of D, L, G and P which occur in the (various *tāla*-forms which are included in the) *n-druta prastāra.* We can easily deduce this fact from the inductive construction of *prastāra* outlined earlier. For instance, let D_n be the total number of *drutas* in the *n-druta prastāra.* The last S_{n-1} rows which end in a D have $S_{n-1} + D_{n-1}$ *drutas.* The number of *drutas* in the rows which end in L, G and P are D_{n-2}, D_{n-4} and D_{n-6} respectively. Hence, D_n satisfy the recurrence relation

$$D_n = S_{n-1} + D_{n-1} + D_{n-2} + D_{n-4} + D_{n-6}$$

which is the same as the recurrence relation (20) satisfied by the *pātāla* numbers T_n. The equality of D_n and T_n follows from this and the fact that $D_i = T_i$ for small values of i.

Further

$$M_n = T_n + T_{n-1} + T_{n-3} + T_{n-5} = T_{n+1} - S_n, \qquad (21)$$

gives what is sometimes called the *mahā-pātāla*, which is the total number of all *tālāṅgas* (D, L, G and P) which appear in the *n-druta prastāra*.

In Table 12, we give the *pātāla* sequence. Here, $T_0 = 0$, $T_1 = 1$, $T_2 = 1+1 = 2$, $T_3 = 2+2+1 = 5$, $T_4 = 3+5+2 = 10$, $T_5 = 6+10+5+1 = 22$, $T_6 = 10+22+10+2 = 44$, $T_7 = 19+44+22+5+1 = 91$, and so on.

---

[10] In other words, we set $T_i = 0$, for $i \leq 0$.

n	0	1	2	3	4	5	6	7	8	9	10
S_n	1	1	2	3	6	10	19	33	60	106	169
T_n	0	1	2	5	10	22	44	91	180	358	698
M_n	0	1	3	7	16	34	72	147	298	592	1169

TABLE 12. *Pātāla* sequence

3.12. *Druta-meru*

पङ्क्तिं कृत्वेष्टतालस्थद्रुतसंमितकोष्ठिकाम्॥
तिरश्चीं तत्परामूनां कोष्ठेनाथ ततः पराः ।
द्विद्विकोष्ठोनिताः स्वस्वपूर्वतोऽथाङ्कयोजना ॥
द्वौ द्वौ तासामाद्यकोष्ठौ स्यातामेकाङ्कसंयुतौ॥
अधस्तन्यास्तृतीयादौ विषमे कोष्ठके लिखेत्॥
अन्त्याद्यङ्कचतुष्कस्य योगं संख्याङ्कसंघवत्।
समे त्वन्यं विनैतेषां योगं न्यस्याथ पङ्क्तिषु ॥
परासु शेषकोष्ठेषु चतुर्योगोऽयमिष्यते ।
नास्ति प्रतिनिधिस्त्वासामङ्कयोस्तुर्यषष्ठयोः॥
तासु स्वभावतो यास्तु निष्पन्ना ऊर्ध्वपङ्क्तयः ।
तन्मध्ये तु समा कोष्ठेऽन्त्यस्थानेऽन्त्यादधस्तनः॥
अङ्कः कार्योऽथ तैरङ्कैर्भेदसंख्याभिधीयते ।
विषमायामूर्ध्वपङ्क्तौ स्थितैरङ्कैरधः क्रमात्॥
एकद्रुताद्या विषमत्र्यादिसंख्या द्रुता भिदाः।
सर्वद्रुतान्ता मीयन्ते समपङ्क्तिस्थितैः पुनः ॥
द्रुतहीनादयो द्व्यादिसमसंख्या द्रुता भिदाः।
सर्वद्रुतान्ता ज्ञायन्ते द्रुतमेरुरयं मतः॥
ऊर्ध्वपङ्क्तिस्थसर्वाङ्कयोगात् संख्यापि गम्यते।

(सङ्गीतरत्नाकरः ५.३३८-३४७)

After considering the *pratyaya pātāla*, Śārṅgadeva begins the discussion of the *pratyayas druta-meru, laghu-meru, guru-meru* and *pluta-meru*, which are analogous to the *pratyaya eka-dvayādi-laga-kriyā* for *mātrā-vṛttas*. There, the *druta-meru* is to be constructed to compute the number of *tāla*-forms in a *prastāra* which have a given number of *drutas*.

While considering an *n-druta prastāra*, start with a first row which has n entries. The row above will not have any entry in the first column, the successive rows after that will have two columns less at each stage. In each row, the first two entries from the left will be 1, 1. In the first row, the odd places are filled with the

							1	1
					1	1	7	8
			1	1	5	6	20	27
	1	1	3	4	9	14	25	44
1	1	2	2	5	4	12	7	26
1	2	3	4	5	6	7	8	9

TABLE 13. *Druta-meru*

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 2, 4, etc., *drutas* respectively when n is even, and the number of *tāla*-forms which have 1, 3, 5, etc., *drutas* respectively when n is odd.

sum of the previous, second- previous, fourth-previous and the sixth-previous entries as in the case of *saṃkhyāṅkas.* The even places are to be filled with the same sum except for the previous entry. Where the fourth or sixth-previous entry is not there, the third or the fifth-previous entry is to be used as the representative. In the other rows, the odd places are filled with the sum of the previous, second-previous, fourth-previous and the sixth-previous entries as in the case of *saṃkhyāṅkas.* The even places are to be filled with the same sum except that the previous entry is to be taken from the row below. Where the fourth or sixth-previous entry is not there, no representative is to be used in their place.

Then, the entries read from below in the n-th column give the number of *tāla*-forms in the n-*druta prastāra* with $1, 3, 5$, etc., *drutas* respectively, in the case when n is odd; they give the number of *tāla*-forms in the n-*druta prastāra* with $0, 2, 4$, etc., *drutas* respectively, in the case when n is even.[11] Notice that the sum of all the entries in any column is always equal to the corresponding *saṃkhyāṅka.*

In Table 13, a bottom (or zeroth) row is introduced below the first row to denote the total value of the *prastāra.* The rows of the *meru* are filled as follows.

- First row odd entries: $1 = 1,\ 2 = 1+1,\ 5 = 2+2+1,\ 12 = 4+5+2+1,\ 26 = 7+12+5+2$
- First row even entries: $1 = 1,\ 2 = 1+1,\ 4 = 2+1+1,\ 7 = 4+2+1$ (in second and third sums 1 is the *pratinidhi* or the representative)
- Second row odd entries: $4 = 3+1,\ 14 = 9+4+1,\ 44 = 25+14+4+1$

[11] Since L, G and P correspond to *kāla-pramāṇas* 2, 4 and 6 which are all even, a n-*druta prastāra* will have rows which contain only odd number of Ds when n is odd, and only even number of Ds when n is even.

- Second row even entries: $3 = 2 + 1,\ 9 = 5 + 3 + 1,\ 25 = 12 + 9 + 3 + 1$

For example, in the 5-*druta prastāra*, there are 5 *tāla*-forms with 1 *druta*, 4 with 3 *drutas*, and 1 with 5 *drutas*. In the 6-*druta prastāra*, there are 4 *tāla*-forms with 0 *drutas*, 9 with 2 *drutas*, 5 with 4 *drutas* and 1 with 6 *drutas*. In the 7-*druta prastāra*, there are 12 *tāla*-forms with 1 *druta*, 14 with 3 *drutas*, 6 with 5 *drutas* and 1 with 7 *drutas*.

The above construction of the *druta-meru* is in fact a generalisation of the construction of the *mātrā-meru* used in the *eka-dvayādi-lagakriyā* for the *mātrā-vṛttas* in the *Prākṛta Paiṅgala* (in Prākṛta, c. 1300) or the commentary *Nārāyaṇī* of Nārāyaṇabhaṭṭa (c.1555) on *Vṛttaratnākara* of Kedārabhaṭṭa. There, the *meru* is constructed essentially by a recurrence relation of the type $g_k^n = g_k^{n-1} + g_{k-1}^{n-2}$, where g_k^n is the number of metrical forms in the *prastāra* of n-*mātrās* with k *gurus*. These can be shown to be related to the binomial coefficients and their sum for any given n will be related to the *saṃkhyāṅkas* or the Virahāṅka-Fibonacci numbers.[12]

Let D_k^n denote the number of *tāla*-forms with k *drutas* in the n-*druta prastāra*. If we refer back to the inductive construction of the *prastāra* explained earlier, we can easily see that of these *tāla*-forms, D_{k-1}^{n-1} will be the number of *tāla*-forms which end in a *druta*, while D_k^{n-2}, D_k^{n-4} and D_k^{n-6} will be the number of *tāla*-forms which end in a *laghu*, *guru* and *pluta*, respectively. Thus we have the recurrence relation:[13]

$$D_k^n = D_{k-1}^{n-1} + D_k^{n-2} + D_k^{n-4} + D_k^{n-6} \qquad (\text{for } n > 6 \text{ and } k > 0) \tag{22}$$

The above recursion relation satisfied by D_k^n, is in fact the basis for the above construction of the *druta-meru*.

We shall now outline an alternate derivation of the above relation (22) using the generating function approach. For this purpose, we first note that the coefficient of $t^k x^n$ in $(tx + x^2 + x^4 + x^6)^p$ gives the number of possible *tāla*-forms of value (*kāla-pramāṇa*) n with p *talāṅgas* (i.e., D, L, G or P) of which k are *drutas* (D). Summing over all the possibilities, we get

$$[1 - (tx + x^2 + x^4 + x^6)]^{-1} = \sum_n \left[\sum_{k<n} D_k^n t^k \right] x^n \tag{23}$$

---

[12]For details, we refer the reader to the mathematical theory of *mātrā-vṛtta-prastāra* outlined in the paper of R. Sridharan, cited earlier.

[13]For $1 \le n \le 6$, or $k = 0$, we only need to take those terms which are meaningful in the recurrence relation (22).

We therefore have

$$[1 - (tx + x^2 + x^4 + x^6)] \sum_n \left[\sum_{k<n} D_k^n t^k \right] x^n = 1 \tag{24}$$

Comparing the coefficients of $t^k x^n$ on both sides of the above equation, we obtain the recurrence relation (22). Incidentally, by setting $t = 1$ in (23), we obtain the obvious relation

$$\sum_{k<n} D_k^n = S_n \tag{25}$$

The recurrence relations (22) are not very transparent in the form in which *druta-meru* of Śārṅgadeva is laid out in Table 13, because the same row corresponds to k or $(k+1)$ *drutas* in alternate columns. In Appendix C, we shall display an alternative form of *druta-meru* proposed by Akella Mallikarjuna Sarma, which clearly reflects the above recurrence relation.

3.13. *Laghu-meru*

लघुमेरौ कोष्ठपङ्क्तीः प्राग्वन्न्यस्येत् तदादिमः ॥
एकैककोष्ठ एकाङ्कयुक्तोऽधः पङ्क्तिकेषु तु ।
शेषकोष्ठेष्वन्त्यतुर्यषष्ठयोगं निवेशयेत् ॥
परासां शेषकोष्ठेषूपान्त्याधस्तनसंयुतम् ।
त्रियोगमेवमादध्यादत्र च द्रुतमेरुवत् ॥
सदसत्वे प्रतिनिधेः कोष्ठाङ्कैरूर्ध्वपङ्क्तिगैः ।
लघुहीनादुपक्रम्यैकाद्येकोत्तर वृद्धलाः ॥
सर्वलान्ताः क्रमाज्ज्येयाः संख्या सर्वाङ्कसंगतेः ।

(सङ्गीतरत्नाकरः: ५.३४७-३५१)

The *laghu-meru* is to be constructed to compute the number of *tāla*-forms in a *prastāra* which have a given number of *laghus*.

Like in the *druta-meru*, while considering an *n-druta prastāra*, start with a row which has n entries. The row above will not have any entry in the first column, the successive rows after that will have two columns less at each stage. In each row, the first entry from the left will be 1. In the first row, the places are filled with the sum of the previous, the fourth-previous and the sixth-previous entries in the same row. Where the fourth or sixth-previous entry is not there, the third or the fifth-previous entry is to be used as the representative. In the other rows, the places are filled with the sum of the previous, fourth-previous and the sixth-previous entries in the same row, and the second-previous entry in the row below.

									1
							1	5	15
					1	4	10	20	39
			1	3	6	10	18	33	61
	1	2	3	4	7	12	21	34	54
1	1	1	2	3	5	7	10	14	21
1	2	3	4	5	6	7	8	9	10

TABLE 14. *Laghu-meru*

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 1, 2 etc., *laghus* respectively.

Where the fourth or sixth-previous entry is not there, no representative is to be used in their place.

Then, the entries read from the bottom in the n-th column give the number of *tāla*-forms in the n-*druta prastāra* with $0, 1, 2, 3$, etc., *laghus* respectively. The sum of all the entries in any column is always equal to the corresponding *saṃkhyāṅkas*.

In Table 14, a bottom (or zeroth) row is introduced below the first row to denote the total value of the *prastāra*. The rows of the *meru* are filled as follows.

First row: $1 = 1,\ 1 = 1,\ 1 = 1,\ 2 = 1 + 1,\ 3 = 2 + 1,\ 5 = 3 + 1 + 1,\ 7 = 5 + 1 + 1,\ 10 = 7 + 2 + 1,\ 14 = 10 + 3 + 1,\ 21 = 14 + 5 + 2$ (in the fourth and sixth entry, 1 is the *pratinidhi*).

Second row: $2 = 1 + 1,\ 3 = 2 + 1,\ 4 = 3 + 1,\ 7 = 4 + 2 + 1,\ 12 = 7 + 3 + 2,\ 21 = 12 + 5 + 3 + 1,\ 34 = 21 + 7 + 4 + 2,\ 54 = 34 + 10 + 7 + 3$.

For example, in the 6-*druta prastāra*, there are 5 *tāla*-forms with 0 *laghus*, 7 with 1 *laghu* 6 with 2 *laghus* and 1 with 3 *laghus*. In the 7-*druta prastāra*, there are 7 *tāla*-forms with 0 *laghus*, 12 with 1 *laghu*, 10 with 2 *laghus* and 4 with 3 *laghus*.

If L_k^n denotes the number of *tāla*-forms with k *laghus* in the n-*druta prastāra*, then we can easily see that we have the following relation analogous to (23)

$$[1 - (x + tx^2 + x^4 + x^6)]^{-1} = \sum_n \left[\sum_{k<n} L_k^n t^k \right] x^n \qquad (26)$$

From (26) we obtain the following recurrence relation

$$L_k^n = L_k^{n-1} + L_{k-1}^{n-2} + L_k^{n-4} + L_k^{n-6} \text{ (for } n > 6 \text{ and } k > 0), \qquad (27)$$

which is clearly the basis for the construction of the *laghu-meru* as described above.

3.14. *Guru-meru*

गुरुमेरावधः पङ्क्तेः परा कोष्ठत्रयोनिता ॥
चतुश्चतुष्कोष्ठहीनाः स्वस्वपूर्वावलेः पराः ।
एकाङ्कवन्त आद्याद्यकोष्ठाः प्रथमपङ्क्तिगः ॥
द्वितीयो द्व्यङ्कवानन्त्योपान्त्यषष्ठाङ्कयोगिनः ।
शेषकोष्ठाः परासां तु द्वितीयादिषु लिख्यते ॥
योगोऽन्त्योपान्त्यषष्ठानामधस्तुर्याङ्कसंयुतः ।
लघुस्थाने गुरुर्ज्ञेयः शेषं तु लघुमेरुवत् ॥

(सङ्गीतरत्नाकरः ५.३५१-३५४)

The *guru-meru* is to be constructed to compute the number of *tāla* forms in a *prastāra* which have a given number of *gurus*.

While considering an *n-druta prastāra*, start with a row which has n entries. The row above will have three columns less and the subsequent rows will have four columns less at each stage. In each row, the first entry from the left will be 1. In the first row, place 2 at the second place. The other places are filled with the sum of the previous, second-previous and the sixth-previous entries. When the sixth-previous entry is not there, the fifth-previous entry is to be used as the representative. In the other rows, the places are filled with the sum of the previous, second-previous and the sixth-previous entries in the same row, and the fourth-previous entry in the row below. Where the sixth-previous entry is not there, no representative is to be used in its place.

Then, the entries read from the bottom in the n-th column give the number of *tāla*-forms in the *n-druta prastāra* with 0, 1, 2, 3, etc., *gurus* respectively. The sum of all the entries in any column is always equal to the corresponding *saṃkhyāṅka*.

In Table 15, a bottom (or zeroth) row is introduced below the first row to denote the total value of the *prastāra*. The rows of the *meru* are filled as follows:

First row: $1 = 1,\ 2 = 2,\ 3 = 2 + 1,\ 5 = 3 + 2,\ 8 = 5 + 3,\ 14 = 8+5+1,\ 23 = 14+8+1,\ 39 = 23+14+2,\ 65 = 39+23+3,\ 109 = 65+39+5$ (in the sixth entry 1 is the *pratinidhi*).

Second row: $2 = 1+1,\ 5 = 2+1+2,\ 10 = 5+2+3,\ 20 = 10+5+5,\ 38 = 20 + 10 + 8,\ 73 = 38 + 20 + 14 + 1$.

For example, in the 6-*druta prastāra*, there are 14 *tāla*-forms with 0 *gurus* and 5 with 1 *guru*. In the 9-*druta prastāra*, there are 65 *tāla*-forms with 0 *gurus*, 38 with 1 *guru* and 3 with 2 *gurus*.

							1	3	9
			1	2	5	10	20	38	73
1	2	3	5	8	14	23	39	65	109
1	2	3	4	5	6	7	8	9	10

TABLE 15. *Guru-meru*

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 1, 2 etc., *gurus* respectively.

If G_k^n denotes the number of *tāla*-forms with k *laghus* in the n-*druta prastāra*, then we can easily see that we have the following relation analogous to (23):

$$[1-(x+x^2+tx^4+x^6)]^{-1} = \sum_n \left[\sum_{k<n} G_k^n t^k\right] x^n \qquad (28)$$

From (28) we obtain the following recurrence relation

$$G_k^n = G_k^{n-1} + G_k^{n-2} + G_{k-1}^{n-4} + G_k^{n-6} \qquad (\text{for } n > 6 \text{ and } k > 0), \qquad (29)$$

which is clearly the basis for the construction of the *guru-meru* as described above.

3.15. *Pluta-meru*

प्लुतमेरावधःपङ्क्तेः पञ्चकोष्ठोनिता परा ।
तत्पराः स्वस्वपूर्वातः षड्वङ्कोष्ठोनिता मताः ॥
अधःपङ्क्तौ तु षष्ठाङ्कस्थाने तुर्यं नियोजयेत् ।
त्रियोगोऽयमऽधः षष्ठयुक्तः स्यात् परपङ्क्तिषु ॥
प्लुतो ज्ञेयो गुरुस्थाने शेषं तु गुरुमेरुवत्।

(सङ्गीतरत्नाकरः ५.३५४–३५७)

The *pluta-meru* is to be constructed to compute the number of *tāla* forms in a *prastāra* which have a given number of *plutas*.

While considering an n-*druta prastāra*, start with a row which has n entries. The row above will have five columns less and the subsequent rows will have six columns less at each stage. In each row, the first entry from the left will be 1. In the first row, place 2 at the second place. The other places are filled with the sum of the previous, second-previous and the fourth-previous entries. When

the fourth-previous entry is not there, the third-previous entry is to be used as the representative. In the other rows, the places are filled with the sum of the previous, second-previous, and the fourth-previous entries in the same row, and the sixth-previous entry in the row below. Where the fourth-previous entry is not there, no representative is to be used in its place. The sum of all the entries in any column is always equal to the corresponding *saṃkhyāṅka*.

					1	2	5	10	22	44
1	2	3	6	10	18	31	55	96	169	296
1	2	3	4	5	6	7	8	9	10	11

TABLE 16. *Pluta-meru*

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 1, 2 etc., *plutas* respectively.

In Table 16, a bottom (or zeroth) row is introduced below the first row to denote the total value of the *prastāra*. The rows of the *meru* are filled as follows.

First row: $1 = 1,\ 2 = 2,\ 3 = 2 + 1,\ 6 = 3 + 2 + 1,\ 10 = 6 + 3 + 1,\ 18 = 10 + 6 + 2,\ 31 = 18 + 10 + 3,\ 55 = 31 + 18 + 6,\ 96 = 55 + 31 + 10,\ 169 = 96 + 55 + 18,\ 296 = 169 + 96 + 31$ (in the fourth entry 1 is the *pratinidhi*).

Second row: $2 = 1 + 1,\ 5 = 2 + 1 + 2,\ 10 = 5 + 2 + 3,\ 22 = 10 + 5 + 1 + 6,\ 44 = 22 + 10 + 2 + 10$.

For example, in the 6-*druta prastāra*, there are 18 *tāla*-forms with 0 *plutas* and 1 with 1 *pluta*. In the 9-*druta prastāra*, there are 96 *tāla*-forms with 0 *plutas*, and 10 with 1 *pluta*.

If P_k^n denotes the number of *tāla*-forms with k *plutas* in the n-*druta prastāra*, then we can easily see that we have the following relation analogous to (23):

$$[1 - (x + x^2 + x^4 + tx^6)]^{-1} = \sum_n \left[\sum_{k<n} P_k^n t^k \right] x^n \qquad (30)$$

From (30) we obtain the following recurrence relation

$$P_k^n = P_k^{n-1} + P_k^{n-2} + P_k^{n-4} + P_{k-1}^{n-6} \qquad (\text{for } n > 6 \text{ and } k > 0), \qquad (31)$$

which is clearly the basis for the construction of the *pluta-meru* as described above.

3.16. Other *Pratyayas*

After the discussion of the *druta*, *laghu*, *guru* and *pluta-merus*, Śārṅgadevā goes on to consider ten more *pratyayas*, namely the *saṃyoga-meru*, *khaṇḍā-prastāra* and the *naṣṭa* and *uddiṣṭa* processes associated with the *druta*, *laghu*, *guru* and *pluta-merus*. We shall here merely indicate the nature of these *pratyayas*, which have been dealt with in considerable detail by Śārṅgadevā in over fifty verses (verses 357-409) of Chapter V of *Saṅgītaratnākara*.

The *saṃyoga-meru* gives the number of *tāla*-forms in a *prastāra* which have only D, or only L, or only G, or only P, or only D and L, or only D and G, and so on. This involves a consideration of all the 15 such combinations involving D, L, G and P—taken individually, in pairs, in triples and all the four together.

The next *pratyaya*, *khaṇḍa-prastāra*, is in fact a sub-*prastāra* of a given n-*druta prastāra*, and has only those *tāla*-forms which conform to given specification, such as having no *drutas*, or one *druta*, or two *drutas*, etc., and similarly for *laghus*, *gurus* and *plutas*. Finally, the processes of *naṣṭa* and *uddiṣṭa*, for such a *khaṇḍa-prastāra*, are discussed by Śārṅgadeva in terms of the corresponding *merus*.

4. Acknowledgements

We are thankful to Prof. David Mumford who mooted the idea with Prof. C. S. Seshadri of having a weekly Seminar on History of Mathematics during January-February 2008 at the Chennai Mathematical Institute (CMI). We are indebted to Prof. C. S. Seshadri who, apart from being responsible for making such a Seminar a reality, had the idea of publishing a volume based on the Seminar, to which we are very happy to contribute this article, which has its beginnings in a lecture given in the Seminar by one of the authors (R. Sridharan). We are grateful to the referees of this paper for their valuable comments and suggestions. We are grateful to Prof. K. Ramasubramanian for his kind and enthusiastic help in the preparation of the final manuscript. Raja Sridharan would also like to thank the Institute of Mathematical Sciences Chennai and Prof. R. Balasubramanian, in particular, for having invited him in January 2008, during which visit he was able to attend the Seminar on History of Mathematics at CMI and learn many things.

Appendices

Appendix A. Canonical decomposition and the rule of enumeration of *Tāla-prastāra*

In this Appendix, we shall explain how Śārṅgadeva's rule of enumeration of the *tāla- prastāra* has an elegant mathematical description if viewed in terms of what we have referred to as the canonical decomposition of a row number as a sum of the *saṃkhyāṅkas* S_n or Śārṅgadeva numbers. We have already noted earlier (in our discussion of the *naṣṭa* process) that the m-th row in the n-*druta-prastāra* is obtained by taking the canonical decomposition of $S_n - m$, writing 1 for all the *patita* numbers ($< S_n$) and 0 for the others, substituting P for the string 010110, G for 0110, L for 10 and D for the string consisting of the single symbol 0. The next lemma shows that for $m = S_n - 1$, this actually leads to the first row of the *prastāra* as described by Śārṅgadeva.

Lemma: The canonical decomposition of $S_n - 1$ leads via the process of *naṣṭa* to the first row of the n-*druta-prastāra*, which has the maximum number of Ps at the extreme right followed by a DG or G or DL or L or a D on the left, depending on whether $n = 1, 2, 3, 4, 5$ (modulo 6). The first row has only Ps if $n = 0$ (modulo 6).

Proof: If $n = 1$, $S_n - 1 = 0$ and the first row of the 1-*druta-prastāra* is just D. If $n = 2$, $S_n - 1 = 1 = 1.S_1 + 0.S_2$ and hence the first row of the 2-*druta-prastāra* corresponds to the sequence 10 and is therefore an L. If $n = 3$, $S_n - 1 = 2 = 0.S_1 + 1.S_2 + 0.S_3$, hence the first row of the 3-*druta-prastāra* corresponds to the sequence 010 and is hence DL. If $n = 4$, $S_n - 1 = 6 - 1 = 5 = 0.S_1 + 1.S_2 + 1.S_3 + 0.S_4$, hence the first row of the 4-*druta-prastāra* corresponds to the sequence 0110 and is hence G. If $n = 5$, $S_n - 1 = 10 - 1 = 9 = 0.S_1 + 0.S_2 + 1.S_3 + 1.S_4 + 0.S_5$, hence the first row of the 5-*druta-prastāra* corresponds to the sequence 00110 or DG. If $n = 6$, $S_n - 1 = 19 - 1 = 18 = 0.S_1 + 1.S_2 + 0.S_3 + 1.S_4 + 1.S_5 + 0.S_6$, hence the first row of the 6-*druta-prastāra* corresponds to the sequence 010110 and is hence P.

Let now $n > 6$. We have

$$\begin{aligned} S_n &= S_{n-6} + S_{n-4} + S_{n-2} + S_{n-1} \\ \text{so that} \quad S_n - 1 &= S_{n-6} - 1 + S_{n-4} + S_{n-2} + S_{n-1} \end{aligned} \qquad (32)$$

with $n - 6 > 0$. We may assume by induction on n that the canonical decomposition for $S_{n-6} - 1$ leads to the first row of the $(n - 6)$-*druta-prastāra*. In view of the above equation, the row corresponding to the canonical decomposition of

$S_n - 1$ is obtained by adding only one more P to the right of the first row of the $(n-6)$-*druta-prastāra* and our lemma is thus proved.[14]

We shall now see that the above lemma can be used to relate the rule of Śārṅgadeva for passing from the m-th row of the n-*druta-prastāra* to the $(m+1)$-th row, via the change in the canonical decomposition of $S_n - m$ to that of $S_n - (m+1)$.

Suppose that the first non-*druta* element (from the left) of the m-th row is a P. Then the canonical decomposition of $S_n - m$ has the binary form $00\ldots010110\ldots$. Let us assume that the first 1 in the above string occurs at the k-th place (counted from the left), i.e., that the canonical decomposition of $S_n - m$ begins with a S_k. Therefore, in the canonical decomposition of $S_n - (m+1)$, the 1 in the k-th position in the canonical decomposition of $S_n - m$ changes to a zero and, in order to obtain the canonical decomposition of $S_n - (m+1)$, we need to know the canonical decomposition of $S_k - 1$, which has been obtained in the above lemma. Thus we see that while going from the m-th to $(m+1)$-th row, the P (above) changes to a G and we have to insert, from right to left, as many Ps as possible followed by DG, or G or DL or L or D or none of them as the case may be. The symbols to the right of P in the m-th row remain unchanged in the $(m+1)$-th row. This is precisely the rule given by Śārṅgadeva for going from a given row to the next in the *prastāra*, when the first non-*druta* element in the row (going from left to right) happens to be P. If the first non-*druta* element in the m-th row of a *prastāra* happens to be G or L, we can use a similar argument to show that in the next row this symbol is changed to L or D respectively, and the rest of the symbols of the row on the left are obtained exactly as given by the rule of Śārṅgadeva, and those on the right remain the same.

Let us illustrate this by an example. The 15-th row (from the top) of the 7-*druta prastāra* with 33 rows (see Table 6) is PD and corresponds to the canonical decomposition

$$\begin{aligned} S_7 - 15 &= 33 - 15 = 18 = 2 + 6 + 10 \\ &= 0.S_1 + 1.S_2 + 0.S_3 + 1.S_4 + 1.S_5 + 0.S_6 + 0.S_7 \end{aligned}$$

[14] We may note that, in the case of *mātrā-vṛttas*, the above lemma which gives the first row of a *prastāra* of value n translates itself into the following simple identities between Virahāṅka-Fibonacci numbers (see the article of R. Sridharan, cited earlier, p. 133):

$$\begin{aligned} s_1 + s_3 + \ldots + s_{2k-1} &= s_{2k} - 1 \quad (\text{if } n = 2k) \\ s_2 + s_4 + \ldots + s_{2k} &= s_{2k+1} - 1 \quad (\text{if } n = 2k+1) \end{aligned}$$

This corresponds to the 7-tuple $(0, 1, 0, 1, 1, 0, 0)$. The next row is associated with the canonical decomposition

$$\begin{aligned} S_7 - 16 &= 33 - 16 = 17 = 1 + 6 + 10 \\ &= (S_2 - 1) + 0.S_3 + 1.S_4 + 1.S_5 + 0.S_6 + 0.S_7 \\ &= 1.S_1 + 0.S_2 + 0.S_3 + 1.S_4 + 1.S_5 + 0.S_6 + 0.S_7 \end{aligned}$$

This corresponds to the 7-tuple $(1, 0, 0, 1, 1, 0, 0)$ and represents the row LGD (see Table 6). That is, P is replaced by LG and the symbol D to the right of P remains unchanged.

In Table 17, we have listed the 33 binary sequences of length 7 that correspond to the rows of the 7-*druta-prastāra*. We note that the last row whose entries are all 0 corresponds to the canonical decomposition of 0. The row above it is $(1, 0, ..., 0)$ is got by increasing the first zero of the last row by 1. This corresponds to replacing $S_1 - 1 = 0$ by $S_1 = 1$. In fact, in passing from the m-th row to the $(m + 1)$-th row from below, a certain zero entry becomes 1 and the rest of the entries to the left become 0 and entries to the right are unchanged. This, indeed, corresponds to replacing the canonical decomposition of $S_k - 1$ for some $k < n$ (in a n-*druta prastāra*), by S_k.

For instance, in Table 17 below (for $n = 7$), the passage from 27th row (counted from below) which corresponds to the binary sequence $(0, 1, 0, 1, 0, 1, 0)$ to the 28th, the sequence in question gets replaced by $(0, 0, 1, 1, 0, 1, 0)$ which corresponds to the replacement of $S_3 - 1$ by S_3. In the same way, the passing from the 15th row (counted from below) which corresponds to the binary sequence $(0, 1, 1, 0, 1, 0, 0)$ to the 16th, the above sequence gets replaced by $(0, 0, 0, 1, 1, 0, 0)$, which corresponds in turn to the replacement of $S_4 - 1$ by S_4.

It is interesting to note that the process is analogous to the rule for going from one row to the next in the case of *tāna prastāra*, where it is discussed in terms of the factorial representation (see Proposition 2, Sec. 2).

We shall now indicate a method of constructing the binary sequences corresponding to the rows of the *prastāra* of value n, assuming that such a description is possible for the *prastāra* of value $n - 1$. To do this, we first show how the n-*druta prastāra* itself can be obtained through the knowledge of the $(n - 1)$-*druta prastāra*.

As we have remarked earlier, the *prastāra* of value n is obtained by first writing down the *prastāra* of value $n - 6$, augmented by P, adding below this the *prastāra* of value $n - 4$ augmented by G, adding below this the *prastāra* of value $n - 2$ augmented by L and finally adding below this the *prastāra* of value $n - 1$ augmented by D. We will see how the *prastāras* of values $n - 2$, $n - 4$ and $n - 6$ sit inside the *prastāra* of value $n - 1$. If we take the S_{n-2} rows of the *prastāra* of value $n - 1$ ending with D and omit the last D, we obtain the *prastāra* of value

m									S_1 (1)	S_2 (2)	S_3 (3)	S_4 (6)	S_5 (10)	S_6 (19)	S_7 (33)	S_7-m (33-m)
1						D	P		0	0	1	0	1	1	0	32
2					D	L	G		0	1	0	0	1	1	0	31
3					L	D	G		1	0	0	0	1	1	0	30
4				D	D	D	G		0	0	0	0	1	1	0	29
5					D	G	L		0	0	1	1	0	1	0	28
6				D	L	L	L		0	1	0	1	0	1	0	27
7				L	D	L	L		1	0	0	1	0	1	0	26
8			D	D	D	L	L		0	0	0	1	0	1	0	25
9					G	D	L		0	1	1	0	0	1	0	24
10				L	L	D	L		1	0	1	0	0	1	0	23
11			D	D	L	D	L		0	0	1	0	0	1	0	22
12			D	L	D	D	L		0	1	0	0	0	1	0	21
13			L	D	D	D	L		1	0	0	0	0	1	0	20
14		D	D	D	D	D	L		0	0	0	0	0	1	0	19
15						P	D		0	1	0	1	1	0	0	18
16					L	G	D		1	0	0	1	1	0	0	17
17				D	D	G	D		0	0	0	1	1	0	0	16
18					G	L	D		0	1	1	0	1	0	0	15
19				L	L	L	D		1	0	1	0	1	0	0	14
20			D	D	L	L	D		0	0	1	0	1	0	0	13
21			D	L	D	L	D		0	1	0	0	1	0	0	12
22			L	D	D	L	D		1	0	0	0	1	0	0	11
23		D	D	D	D	L	D		0	0	0	0	1	0	0	10
24				D	G	D	D		0	0	1	1	0	0	0	9
25			D	L	L	D	D		0	1	0	1	0	0	0	8
26			L	D	L	D	D		1	0	0	1	0	0	0	7
27		D	D	D	L	D	D		0	0	0	1	0	0	0	6
28				G	D	D	D		0	1	1	0	0	0	0	5
29			L	L	D	D	D		1	0	1	0	0	0	0	4
30		D	D	L	D	D	D		0	0	1	0	0	0	0	3
31		D	L	D	D	D	D		0	1	0	0	0	0	0	2
32		L	D	D	D	D	D		1	0	0	0	0	0	0	1
33	D	D	D	D	D	D	D		0	0	0	0	0	0	0	0

TABLE 17. Canonical decomposition of rows of Seven-*druta-prastāra*

$n-2$. If we take the S_{n-4} rows of the *prastāra* of value $n-1$ ending with DL and omit the last two entries DL from the rows, we obtain the *prastāra* of value $n-4$. This is because, if we take the rows of the *prastāra* of value $n-1$ ending with L and omit the last entry L, we obtain the *prastra* of value $n-3$; now, if we take the rows of the *prastāra* of value $n-3$ ending with a D and omit this D, we obtain the *prastāra* of value $n-4$. It follows by a similar argument that if we take the S_{n-6} rows of the *prastāra* of value $n-1$ ending with DLL and delete these three entries from the rows, we obtain the *prastāra* of value $n-6$. From the above

remarks, we deduce the following:

Proposition: The *prastāra* of value n is obtained from the *prastāra* of value $n-1$ as follows:

(i) We take all the rows of the *prastāra* of value $n-1$ and augment these by a D at the end. These form the last S_{n-1} rows of the *prastāra* of value n.

(ii) We take the rows of the *prastāra* of value $n-1$ ending with a D and replace this D by an L. These form the next S_{n-2} rows of the *prastāra* of value n occurring above the rows constructed in (i).

(iii) We take the rows of the *prastāra* of value $n-1$ ending with DL and replace the last two entries DL of these rows by a G. These form the next S_{n-4} rows of the *prastāra* of value n occurring above the rows constructed in (ii).

(iv) Finally, we take the rows of the *prastāra* of value $n-1$ ending with DLL and replace the last three entries DLL of these rows by a P. We then obtain the first S_{n-6} rows of the *prastāra* of value n.

In a similar vein, we can construct the binary sequences of length n associated with the *prastāra* of value n, assuming that we have already constructed the binary sequences of length $n-1$ associated with the *prastāra* of value $n-1$. To carry out this construction, we need to write down the canonical decomposition of any number less than S_n assuming that we have written down the canonical decomposition of any number less than S_{n-1}. Now, any integer less than S_n is either less than S_{n-1} or of the form $S_{n-1}+i$ where $0 \le i < S_n - S_{n-1}$. Since $2S_{n-1} \ge S_n$, we have $i < S_{n-1}$. Therefore we can write down the canonical decomposition of any number less than S_n (assuming that we have written down the canonical decomposition of any number less than S_{n-1}). This leads to the following procedure for the construction of the binary sequences of length n that we are looking for.

In order to write down the binary sequences of length n associated with the *prastāra* of value n:

(i) We take all the binary sequences of length $n-1$ corresponding to the *prastāra* of value $n-1$ and augment these sequences on the right by 0. These form the binary sequences corresponding to the last S_{n-1} rows of the *prastāra* of value n. This amounts to augmenting each of the rows of the *prastāra* of value $n-1$ by D.

(ii) The binary sequences associated with the first $S_n - S_{n-1}$ rows of the *prastāra* of value n are obtained as follows: We take the binary sequences associated with the last $S_n - S_{n-1}$ rows of the *prastāra* of value $n-1$ and replace the last 0 of these rows by the string 10. We can verify that this procedure amounts to taking the rows of the *prastāra* of value $n-1$ ending with a D and replacing this D by an L, taking the rows of the *prastāra* ending with

1							
2	1						
	2	1					
4		2	1				
	4		2	1			
6		4		2	1		
	6		4		2	1	
1	1	2	3	6	10	19	33

TABLE 18. *Unmeru* for *Tāla Prastāra*

DL and replacing the DL by G and finally taking the rows of the *prastāra* ending with DLL and replacing the DLL with a P.

Remark: It is interesting to see what happens if we apply the procedure outlined in (ii) to all the binary sequences associated with the *prastāra* of value $n-1$. For example, if $n = 8$, $S_8 - S_7 = 60 - 33 = 27$. We have already seen in (ii) what happens if we take the last 27 rows of the 7-*druta prastāra* and replace the last 0 of these sequences with the string 10. Now, if we take binary strings associated with the first six rows of the 7-*druta prastāra* and replace the last 0 of these sequences with the string 10, then we see that in each case one of the conditions (a)–(e) (mentioned in Proposition 3 of Sec. 3) will be violated.

Appendix B. Nārāyaṇa's *unmeru* method for *naṣṭa* and *uddiṣṭa*

Nārāyaṇa Paṇḍita in his *Gaṇita-kaumudī* (c. 1356) discusses the *naṣṭa* and *uddiṣṭa* processes for his generalization of the *mātrā-vṛtta-prastāra*. For this purpose, he introduces what he refers to as the *unmeru*. Though Nārāyaṇa Paṇḍita's generalization of the *mātrā-vṛtta-prastāra* does not subsume the *tāla-prastāra* as considered in *Saṅgītaratnākara*, his *unmeru* method can be suitably altered so as to give an alternative method for *naṣṭa* and *uddiṣṭa*.

We shall illustrate the modified *unmeru* method for *naṣṭa* and *uddiṣṭa* by considering the example of the 7-*druta prastāra*. To construct the *unmeru*, in the bottom row place the number 1, followed by the *saṃkhyāṅkas* up to 33, the *saṃkhyāṅka* of the given *prastāra*. In the column above the *saṃkhyāṅkas*, place the integers, 1, 2, 3, 4, etc., starting with one entry above the penultimate *saṃkhyāṅka* 19, with the number of entries in each subsequent column increasing by one, as we move from right to left. Then, remove all entries in the columns above the *saṃkhyāṅkas*, except the values 1, 2, 4, 6 corresponding to our four *tālāṅgas*. Thus the *unmeru* is formed as in Table 18.

B.1. *Naṣṭa-vidhi*

Given any number less than or equal to 33, find the set of *patita* and *apatita saṃkhyāṅkas* as usual. Starting from the right, note the entry in the first row above the first *apatita* (if there is no entry go to the next *apatita* along the same row). In the row above the top most entry of the column of that *apatita*, move left till you reach the column of the next *apatita*. Note the corresponding entry (or move on in the same row till the next *apatita*, if there is no entry). And so on.

Ex. 1: To find the 8^{th} *tāla*-form in the 7-*druta prastāra*.

1							
2	**1**						
	2	**1**					
4		2	1				
	4		**2**	1			
6		4		2	1		
	6		4		**2**	1	
1	**1**	**2**	**3**	6	**10**	19	**33**

The *apatita* are 33, 10, 3, 2, 1, 1. We get the *tāla*-form 11122 or $DDDLL$.

Ex. 2: To find the 15^{th} *tāla*-form in the 7-*druta prastāra*.

1							
2	1						
	2	1					
4		2	1				
	4		2	1			
6		4		2	1		
	6		4		2	**1**	
1	**1**	2	**3**	6	10	**19**	**33**

33, 19, 3, 1, 1 are *apatita*. We get the *tāla*-form 61 or PL.

Ex. 3: To find the 28^{th} *tāla*-form in the 7-*druta prastāra*.

1							
2	1						
	2	1					
4		2	1				
	4		2	**1**			
6		4		2	**1**		
	6		4		2	**1**	
1	**1**	2	3	**6**	**10**	**19**	**33**

33, 19, 10, 6, 1, 1 are *apatita*. We get the *tāla*-form 4111 or $GDDD$.

Ex. 4: To find the 2^{nd} *tāla*-form in the 7-*druta prastāra*.

1							
2	1						
	2	1					
4		2	1				
	4		2	1			
6		4		2	1		
	6		4		2	1	
1	1	2	3	6	10	19	33

33, 6, 3, 1, 1 are *apatita*. We get the *tāla*-form 124 or DLG.

Thus, we see that the *naṣṭa*-process becomes somewhat simpler when we make use of the *unmeru*. Since the *uddiṣṭa*-process is quite simple in the standard formulation itself, there is no need to discuss the process in terms of the *unmeru*.

Appendix C. *Druta-meru* proposed by Akella Mallikarjuna Sarma

The *druta-meru* described in the *Saṅgītaratnākara* and displayed explicitly in commentaries, does not render transparent the recursion relation

$$D_k^n = D_{k-1}^{n-1} + D_k^{n-2} + D_k^{n-4} + D_k^{n-6}$$

satisfied by D_k^n (the number of *tāla*-forms with k-*drutas* in the n-*druta-prastāra*). As is clear by its construction, the successive columns of the *druta-meru* correspond to 1-*druta*, 2-*druta* etc., *prastāras*. But, as we noted earlier, the successive rows have a different interpretation, depending on whether we are considering odd or even columns. The row above the bottom row alternatively gives the number of *tāla*-forms with 1 *druta* and 0 *druta*; the next row alternatively gives the number of *tāla*-forms with 3 *drutas* and 2 *drutas*, and so on.

In other words, the *druta-meru* proposed in *Saṅgītaratnākara* is in some sense obtained by fusing together successive rows of a more spread-out *druta-meru* where each row corresponds to a fixed number of *drutas*. This spread-out *druta-meru* will have a whole lot of 0 entries, as there are no *tāla*-forms with even number of *drutas* in a n-*druta prastāra* where n is odd and vice versa. Such a spread out *druta-meru* has been proposed by Akella Mallikarjuna Sarma (see for instance, his book *Saṅgītaranākara: A critical Interpretation*, Hyderabad 2001, p. 26), and is displayed in Table 20. The relation

$$D_k^n = D_{k-1}^{n-1} + D_k^{n-2} + D_k^{n-4} + D_k^{n-6}$$

can easily be seen to be at the basis of the spread-out *druta-meru* (see Table 20), as the indices n, k now correspond to the columns and rows respectively in a one to one manner.

							1	1
					1	1	7	8
			1	1	5	6	20	27
	1	1	3	4	9	14	25	44
1	1	2	2	5	4	12	7	26
1	2	3	4	5	6	7	8	9

TABLE 19. *Druta-meru* of *Saṅgītaratnākara.*

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 2, 4, etc., *drutas* respectively when n is even, and the number of *tāla*-forms which have 1, 3, 5, etc., *drutas* respectively when n is odd.

													1
												1	0
											1	0	12
										1	0	11	0
									1	0	10	0	65
								1	0	9	0	54	0
							1	0	8	0	44	0	200
						1	0	7	0	35	0	147	0
					1	0	6	0	27	0	104	0	357
				1	0	5	0	20	0	70	0	220	0
			1	0	4	0	14	0	44	0	125	0	336
		1	0	3	0	9	0	25	0	63	0	153	0
	1	0	2	0	5	0	12	0	26	0	56	0	118
1	0	1	0	2	0	4	0	7	0	13	0	24	0
0	1	2	3	4	5	6	7	8	9	10	11	12	13

TABLE 20. Alternative form of *Druta-meru* due to Akella Mallikarjuna Sarma.

The different columns correspond to *prastāras* of different values which are indicated in the bottom row. The entries in the successive rows above the bottom row in the n-th column give the number of *tāla*-forms which have 0, 1, 2 etc., *drutas* respectively.

References

Primary Sources: *Chandaḥ-śāstra*

(1) *Chandaḥ-sūtra* of Piṅgalanāga (c. 300 BC): Ed. and Translated into German in A. Weber, Über die Metrik der Inder, in A. Weber, Indische Studien, VIII Band, Harrwitz and Gofsmann, Berlin 1863; Rep. Hilberheim New York, 1973. Ed. with commentary *Mṛtasañjīvanī* of Halāyudha (c. 950), by Viswanatha Sastri, Bibliotheca Indica, Calcutta 1871-1874. Ed. with commentary *Mṛtasañjīvanī* of Halāyudha (c. 950), by Kedaranatha Vasudeva Lakshmana Pansikara, Kavyamala Series No. 91, Bombay 1908, Rep. Chowkhamba 2001.

(2) *Agni-purāṇa*: Ed. Bibliotheca Indica. Ed. by B. Upadhyaya, Chowkhambha, Varanasi 1966. English translation by N. Gangadharan, Vols. 27-30 of Ancient Indian Tradition and Mythology Series, Motilal Banarsidass, Delhi 1987.

(3) *Nārada-purāṇa*: English translation by G. V. Tagare, Vols. 15-19 of Ancient Indian Tradition and Mythology Series, Motilal Banarsidass, Delhi 1981.

(4) *Viṣṇudharmottara-purāṇà*: Ed. Venkateswara Press, Bombay 1912. Ed. by P. Shah, Vol. I Part 3, Gaekwad Oriental Series, Baroda 1958.

(5) *Chandoviciti* of Janāśraya (c. 600): Ed., with a *Vṛtti*, by P. K. Narayana Pillai, Trivandrum 1949.

(6) *Vṛtta-jāti-samuccaya* or *Kaisiṭṭha-chanda* (in Prākṛta) of Virahāṅka (c. 650): Ed. with commentary of Gopāla (c. 1000) by H. D. Velankar, JBBRAS 1929, 1932. Ed., with commentary by Gopala, by H. D. Velankar, Rajasthan Prachya Vidya Pratisthana, Jodhpur 1962.

(7) *Jayadeva-chandas* of Jayadeva (800): Ed. with *Vivṛti* of Harṣaṭa by H. D. Velankar, Haritosamala, Bombay 1949.

(8) *Ratna-mañjūṣā* (c. 800) : Ed., with a commentary by H. D. Velankar, Bharatiya Jnanapith, Kashi 1949.

(9) *Chando'nuśāsana* of Jayakīrti (900): Ed. by H. D. Velankar, Haritosamala, Bombay 1949.

(10) *Chando-ratnākara* of Ratnākara Śānti, Ed by M. Hahn, Nepal Research Center, Kathmandu 1982.

(11) *Mṛtasañjīvanī* of Halāyudha (c. 950) on *Chandaḥ-sūtra* of Piṅgalanāga (c. 300 BC): Ed. by Pt. Kedaranatha Vasudeva Lakshmana Pansikara, Chaukhambha,Varanasi; Rep. 2001.

(12) *Vyākhyā* of Yādavaprakāśa (c. 1000) on *Chandaḥ-sūtra* of Piṅgalanāga (c. 300 BC): Ed. by H. Sinharay, Asiatic Society, Calcutta 1977; Ed. by Sridhara Vasishtha, Nag Publishers, Delhi 2004.

(13) *Vṛtta-ratnākara* of Kedārabhaṭṭa (c. 1000): Ed. with a commentary by S. Rajagopalacharya, Sarasvati Nilaya Press of Vavilla Ramashastri,

Chennai 1888. Ed. by Pansikar, Nirnayasagar Press, Bombay 1908. Ed., with commentaries *Dhīśodhinī* and *Maṇinidhi* and extracts from commentaries *Jñānadīpa* and *Nṛsiṃhabhāgavatīya*, by Chakravarti Iyengar, Mysore 1912. Ed. with *Nārāyaṇī* of Nārāyaṇabhaṭṭa (c. 1555) by Vaidyanatha Sastri Varkala, Kashi Sanskrit Series, Varanasi 1927. Ed. with *Tātparya* of Trivikrama (c. 1100), by Aryendra Sarma it et al, Sanskrit Academy, Osmania University, Hyderabad 1965. Ed., with commentary *Pañjikā* of Rāmacandrakavi (c. 1450) by Silaskandha Mahasthavira, Nirnayasagar Press, Bombay 1948. Ed., with *Nārāyaṇī* of Nārāyaṇabhaṭṭa (c. 1555) and *Setu* of Bhāskarabhaṭṭa (c. 1675) by Madhusudana Sastri, Krishnadasa Academy, Varanasi 1994.

(14) *Chando'nuśāsana* of Hemacandrasūri (c. 1140) with auto-commentary by the author: Published by Devakaran Mulchand, Nirnayasagar Press, Bombay 1912. Ed. by H. D. Velankar, Adhisthata Singhi Jain Sastran Sikshapitha, Bharatiya Vidya Bhavan, Bombay 1961.

(15) *Prākṛta Paiṅgala* (in Prākṛta, c. 1300): Published in Kavyamala, Bombay 1894. Ed with commentary of Viśvanātha Pañcānana, Vaṃśīdhara etc., by Chandra Mohan Ghosh, Bibliotheca Indica, Calcutta 1902. Ed. by B S Vyasa, 2 Vols., Prakrit Text Society, Varanasi, 1959, 1962.

(16) *Kavidarpaṇa*: Ed., with commentary by H. D. Velankar, Annals of BORI 1935. Ed. by H. D. Velankar, Rajasthan Oriental Research Institute, Jodhpur, 1962

(17) *Vāṇī-bhūṣaṇa* of Dāmodara (c. 1550): Published in Kavyamala Series, No 53, Bombay 1895.

(18) *Nārāyaṇī* of Nārāyaṇabhaṭṭa (c. 1555) on *Vṛttaratnākara* of Kedārabhaṭṭa (c. 1000): Ed. by Vaidyanatha Sastri Varkala, Kashi Sanskrit Series, Varanasi 1927. Ed. by Madhusudana Sastri, Krishnadasa Academy, Varanasi 1994.

(19) *Vṛtta-mauktika* of Candraśekhara Bhaṭṭa (c. 1620): Ed. by M. Vinayasagar, Rajasthan Oriental Research Institute, Jodhpur 1965.

(20) *Setu* of Bhāskarabhaṭṭa (c. 1675) on *Vṛttaratnākara* of Kedārabhaṭṭa (c. 1000): Ed. by Madhusudana Sastri, Krishnadasa Academy, Varanasi 1994.

(21) *Vāgvallabha* of Duḥkhabhañjanakavi (c. 1850): Ed., with the commentary of Deviprasadakavi, by Goswami Damodar Sastri, Chowkhambha, Varanasi 1933.

Primary Sources: *Saṅgīta-śāstra*

(1) *Nāṭyaśāstra* of Bharatamuni (c. 200 BC): Ed., with commentary *Abhinavabhāratī* of Abhinavagupta (c. 950), Y. M. R. Kavi, Gaekwad Oriental Series, Baroda 1926. Ed., with commentary *Abhinavabhāratī* of Abhinavagupta (c. 950), by M Shastri 4 Volumes, Benares Hindu

University, Varanasi 1975; Rep. Parimal Publishers, New Delhi 2001. Ed. with Commentary *Abhinavabhāratī* of Abhinavagupta (c. 950) by Parasvanatha Sastri, Sampurnanand Sanskrit University, 3 Volumes, Varanasi 1992-2001.

(2) *Saṅgīta-samayasāra* of Pārśvadeva (c. 1165): Ed. by T. Ganapati Sastri, Trivandrum Sanskrit Series, No. 87, 1925.

(3) *Saṅgīta-ratnākara* of Niśśanka Śārṅgadeva (c. 1225): Ed. with commentary *Kalānidhi* of Kallinātha (c. 1420), 2 Volumes, Anandasram Sanskrit Series, Pune 1896-7. Ed. with *Kalānidhi* of Kallinātha (c. 1420) and *Saṅgīta-sudhākara* of Siṃhabhūpāla (c. 1350), by S. Subramanya Sastri, V. Krishnamacharya and S. Sarada, 4 Vols. Adyar Library, 1943-1986. Chapters 1-4, Ed. with English translation and Notes by Premalata Sharma and R. K. Shringy, 2 Vols. Delhi 1978, 1989.

(4) *Saṅgītopaniṣad-sāroddhāra* of Sudhākarakalaśa (c. 1325): Ed. by U. P. Shah, Gaekwad Oriental Series, Baroda, 1961. Ed. and Translated by Allyn Miner, Indira Gandhi National Centre for Arts, Delhi 1998.

(5) *Nartana-nirṇaya* of Paṇḍarīkaviṭṭhala (c. 1560): Ed. and Translated by R. Sathyanarayana, Indira Gandhi National Centre for Arts, 3 Vols., Delhi 1994.

(6) *Saṅgīta-darpaṇa* of Catura Dāmodara (c. 1600): Ed. by K. Vasudeva Sastri, Thanjavur Saraswati Mahal Series 32, 1952; 2nd Ed. 1989.

(7) *Tālasamudram* (in Tamil) of Vanapāda Cūḍāmaṇi (c. 1650): Ed. by K. Vasudeva Sastri, Thanjavur Saraswati Mahal Series 50, 1955.

(8) *Gāyakalocanam* of Tachur Singaracharya Brothers (in Telugu) 1902.

Primary Sources: *Gaṇita-śāstra*

(1) *Brāhmasphuṭa-siddhānta* of Brahmagupta (c. 628): Mathematics sections translated in H. T. Colebrooke, *Algebra with Arithmetic and Mensuration from the Sanskrit of Brahmagupta and Bhāskara*, London 1817. Ed. by Sudhakara Dvivedi, Medical Hall Press, Varanasi 1902. Ed. by Ramaswarup Sharma, 4 vols., New Delhi 1966.

(2) *Vāsanā* of Pṛthudakasvāmin (c. 850) on *Brāhmasphuṭa-siddhānta* of Brahmagupta (c. 628).

(3) *Gaṇita-sāra-saṅgraha* of Mahāvīra (c. 850): Ed. with English translation by M. Rangacharya, Madras Govt. Press, 1912. Ed. with Hindi translation by L. C. Jain, Jain Samskriti Samrakshaka Sangha, Sholapur 1963.

(4) *Līlāvatī* of Bhāskarācārya (c. 1150): Translated in H. T. Colebrooke, *Algebra with Arithmetic and Mensuration from the Sanskrit of Brahmagupta and Bhāskara*, London 1817. Ed., with *Buddhi-vilāsinī* of Gaṇeśa Daivajña (c. 1545), by V. G. Apte, 2 Vols, Pune 1937. Ed.,

with *Kriyā-kramakarī* of Śaṅkara Vāriyar (c. 1535), by K. V. Sarma, Hoshiarpur 1975.

(5) *Gaṇita-kaumudī* of Nārāyaṇa Paṇḍita (c. 1356): Ed. by Padmakara Dvivedi, 2 Volumes, Varanasi 1936, 1942. Chapters 13-14 edited and translated in T. Kusuba, *Combinatorics and Magic-squares in India: A Study of Nārāyaṇa Paṇḍita's Gaṇita-kaumudī Chapters 13–14*, PhD Dissertation, Brown University 1993. Translated into English by Paramanand Singh, Gaṇita Bhāratī, 1998-2002.

Other Primary Sources

(1) *Bṛhat-saṃhitā* of Varāhamihira (c. 6th century): Ed, with the *Vivṛti* of Bhaṭṭotpala (c. 967), by Sudhakara Dvivedi, 2 vols., Varanasi 1897. Ed, with the *Vivṛti* of Bhaṭṭotpala, by K. C. Dvivedi, 2 vols, Sampurnanand Sanskrit University, Varanasi 1968.

Studies on *Pratyayas*

(1) A. Weber, Über die Metrik der Inder, in A. Weber, Indische Studien, VIII Band, Harrwitz and Gofsmann, Berlin 1863; Rep. Hilberheim New York 1973.

(2) Ludwig Alsdorf, Die Pratyaytas, Ein Beitrag zur Indischen Mathematik, Zeitschrift für Indologie und Iranistik, 9, 1933, 97-157; Translated into English as 'The Pratya-yas: Indian Contribution to Combinatorics', by S. R. Sarma, Indian Journal of History of Science, 26, 1991, 17-61.

(3) D. E. Knuth, *Art of Computer Programming, Vol 2, Seminumerical Algorithms*, Addison Wesley, New York 1969 [Section 4.6.3, refers to Piṅgala's exponention algorithm].

(4) M. Hahn, On the Pratyaya rules of Ratnākara Śānti, Jour. of Oriental Institute Baroda, 30, 1980, p. 61-77.

(5) Parmanand Singh, The So-called Fibonacci Numbers in Ancient and Medieval India, Historia Mathematica, 12, 1985, 229-234.

(6) A. Van Nooten, Binary Numbers in Indian Antiquity, Journal of Indian Philosophy, 21, 1993, 31-50.

(7) T. Kusuba, *Combinatorics and Magic-squares in India: A Study of Nārāyaṇa Paṇḍita's Gaṇita-kaumudī, Chapters 13-14*, PhD Dissertation, Brown University 1993.

(8) Parmanand Singh, The *Gaṇita-kaumudī* of Nārāyaṇa Paṇḍita, Chapter XIII, English Translation with Notes, Gaṇita Bhāratī, 23, 2001, 18-82.

(9) R. Sridharan, Sanskrit prosody, Piṅgala Sūtras and Binary Arithmetic, in G. Emch, M. D. Srinivas and R. Sridharan Eds., *Contributions to the History of Mathematics in India*, Hindustan Book Agency, Delhi 2005, p. 33-62.

(10) D. E. Knuth: *The Art of Computer Programming Vol. IV, Fasc IV, History of Combinatorial Generation*, Addison Wesley New York 2006, p. 61-2.

(11) R. Sridharan, *Pratyayas* for *Mātrāvṛttas* and Fibonacci Numbers, Mathematics Teacher, 42, 2006, 120-137.

Studies on *Tāla-prastāra*

(1) K. Ramachandran, *The Mathematical Basis of Tala System*, S. V. Suppliers, Chennai 1962.

(2) K. Ramachandran, *Tāla Dīpikai* (in Tamil), Chennai 1975; Reprint 2004.

(3) R. Rangaramanuja Ayyangar, *Saṅgītā Ratnākaram: A Study*, Wilco Publishing House, Bombay 1978, p. 235-244.

(4) Subhadra Chaudhuri, *Bhāratīya Saṅgīta mein Tāl aur Rūp-vidhān* (in Hindi), Krishna Brothers, Ajmer 1984.

(5) Akella Mallikarjuna Sarma, *Tālaprastārasāgaram* (in Telugu), Sangita Academy 1985.

(6) Akella Mallikarjuna Sarma, *Indian Genius of Tāla Prastāra*, Sannidhi Sangita Publications, Hyderabad 2001.

(7) Akella Mallikarjuna Sarma, *Tāla Prastāra of Niśśaṅka Śārṅgadeva's Saṅgītaratnākara*, Sannidhi Sangita Publications, Hyderabad 2001.

(8) Akella Mallikarjuna Sarma, *Systematisation of Prastara Details of Desi Talas*, Sannidhi Sangita Publications, Hyderabad 2003.

Raja Sridharan
School of Mathematics
Tata Institute of Fundamental Research, Mumbai 400005
E-mail: sraja@math.tifr.res.in

R. Sridharan
Chennai Mathematical Institute
Plot H1, SIPCOT IT Park, Padur PO, Siruseri 603103
E-mail: rsridhar@cmi.ac.in and

M. D. Srinivas
Centre for Policy Studies
6, Baliah Avenue, Luz, Mylapore, Chennai 600 004
E-mail: mdsrinivas50@gmail.com

What's so Baffling About Negative Numbers? – a Cross-Cultural Comparison

David Mumford

I was flabbergasted when I first read Augustus De Morgan's writings about negative numbers[1]. For example, in the *Penny Cyclopedia* of 1843, to which he contributed many articles, he wrote in the article *Negative and Impossible Quantities*:

> *It is not our intention to follow the earlier algebraists through their different uses of negative numbers. These creations of algebra retained their existence, in the face of the obvious deficiency of rational explanation which characterized every attempt at their theory.*

In fact, he spent much of his life, first showing how equations with these meaningless negative numbers could be reworked so as to assert honest facts involving only positive numbers and, later, working slowly towards a definition of abstract rings and fields, the ideas which he felt were the only way to build a fully satisfactory theory of negative numbers.

On the other hand, every school child today is taught in fourth and fifth grade about negative numbers and how to do arithmetic with them. Somehow, the aversion to these 'irrational creations' has evaporated. Today they are an indispensable part of our education and technology. Is this an example of our civilization advancing since 1843, our standing today on the shoulders of giants and incorporating their insights? Is it reasonable, for example, that calculus was being developed and the foundations of physics being laid — before negative numbers became part of our numerical language!?

The purpose of this article is not to criticize specific mathematicians but first to examine from a cross cultural perspective whether this same order of discovery, the late incorporation of negatives into the number system, was followed in non-Western cultures. Then secondly, I want to look at some of the main figures in

---

[1]De Morgan's attitudes are, of course, well known to historians of Mathematics. But my naïve idea as a research mathematician had been that *at least* from the time of Newton and the Enlightenment an essentially modern idea of real numbers was accepted by all research mathematicians.



Figure 1. Augustus De Morgan

Western mathematics from the late Middle Ages to the Enlightenment and examine to what extent they engaged with negative numbers. De Morgan was not an isolated figure but represents only the last in a long line a great mathematicians in the West who, from a modern perspective, shunned negatives. Thirdly, I want to offer some explanation of why such an air of mystery continued, at least in some quarters, to shroud negative numbers until the mid 19th century. There are several surveys of similar material[2] but, other than describing well this evolution, these authors seem to accept it as inevitable. On the contrary, I would like to propose that the late acceptance of negative numbers in the West was a strange corollary of two facts which were special to the Western context which I will describe in the last section. I am basically a Platonist in believing that there is a single book of mathematical truths that various cultures discover as time goes on. But rather than viewing the History of Mathematics as the unrolling of one God-given linear scroll of mathematical results, it seems to me this book of mathematics can be read in many orders. In the long process of reading, accidents particular to different cultures can result in gaps, areas of math that remain unexplored until well past the time when they would have

---

[2]Three references are (i) Jacques Sesiano, *The Appearance of Negative Solutions in Medieval Mathematics*, Archive for History of the Exact Sciences, vol. 32, pp. 105-150; (ii) Helena Pycior, *Symbols, Impossible Numbers and Geometric Entanglements*, Cambridge Univ. Press, 1997; (iii) Gert Schubring, *Conflicts between Generalization, Rigor and Intuition*, Springer 2005.

been first relevant. I would suggest that the story of negative numbers is a prime example of this effect.[3]

This paper started from work at a seminar at Brown University but was developed extensively at the seminar on the History of Mathematics at the Chennai Mathematical Institute whose papers appear in this volume. I want to thank Professors P. P. Divakaran, K. Ramasubramanian, C. S. Seshadri, R. Sridharan and M. D. Srinivas for valuable conversations and tireless efforts in putting this seminar together. On the US side, I especially want to thank Professor Kim Plofker for a great deal of help in penetrating the Indian material, Professor Jayant Shah for his help with both translations and understanding of the Indian astronomy and Professor Barry Mazur for discussions of Cardano and the discovery of complex numbers. I will begin with a discussion of the different perspectives from which negative numbers and their arithmetic can be understood. Such an analysis is essential if we are to look critically at what early authors said about them and did with them.

1. The Basis of Negative Numbers and Their Arithmetic

It is hard, after a contemporary education, to go back in time to your childhood and realize why negative numbers were a difficult concept to learn. This makes it doubly hard to read historical documents and see why very intelligent people in the past had such trouble dealing with negative numbers. Here is a short preview to try to clarify some of the foundational issues.

Quantities in nature, things we can measure, come in two varieties: those which, by their nature, are always positive and those which can be zero or negative as well as positive, which therefore come in two forms, one canceling the other. When one reads in mathematical works of the past that the writer discards a negative solution, one should bear in mind that this may simply reflect that for the type of variable in that specific problem, negatives make no sense and not conclude that that author believed all negative numbers were meaningless[4]. Below is a table. The first five are ingredients of Euclidean mathematics and the sixth occurs in Euclid (the unsigned case) and Ptolemy (the signed case, labeled as north and south) respectively.

What arithmetic operations can you perform on these quantities? If they are unsigned, then, as in Euclid, we get the usual four operations:

1. $a + b$ OK
2. $a - b$ but only if $a > b$ (as De Morgan insisted so strenuously)

[3] I believe the discovery of Calculus and, especially, simple harmonic motion, the differential equations of sine and cosine, in India and the West provide a second example.

[4] For example, Bhaskara II has a problem in which you must solve for the number of monkeys in some situation, and obviously this cannot be negative.

TABLE I

Modern units	Naturally Positive Quantities	Signed Quantities
positive integer	# of people/monkeys/ apples	
positive real	proportion of 2 lengths (Euclid, Bk V)	
meters	length of movable rigid bar/stick	
meters^2	area of movable rigid flat object	
meters^3	volume of movable rigid object or incompressible fluid	
degrees (of angle)	Measure of a plane angle	distance N/S of equator
dollars		fortune/debt; profit/loss; asset/liability
meters		(a) distance on line/road, rel. to fixed pt, the 'number line' (b) also, height above/below the surface of earth.
seconds		time before or after the present or relative to a fixed event
meters per second		velocity on a line, forwards or backwards
degrees (of temperature)	Kelvin temperature	Fahrenheit or Celsius temperature
grams	Mass or weight of an object	
gram-meters/sec.^{2}		your weight on a scale = force of gravity on your body (a vector)

3. $a * b$ OK but units of the result are different from those of the arguments, e.g. length × length = area, length × length × length = volume

4. a/b OK but again units are different, e.g. length / length = pure number, area / length = length

If they are signed quantities, addition and subtraction are relatively easy – but modern notation obscures how tricky it is to define the actual operation in all cases!

TABLE II

First summand	*Second summand*	*Sum*	*Difference*
a	b	usual $a+b$	$a-b$ if $a>b$ (neg)$(b-a)$ if $b>a$
(neg)a	(neg)b	(neg)$(a+b)$	$b-a$ if $b>a$ (neg)$(a-b)$ if $a>b$
a	(neg)b	$a-b$ if $a>b$ (neg)$(b-a)$ if $b>a$	$a+b$
(neg)a	b	$b-a$ if $b>a$ (neg)$(a-b)$ if $a>b$	(neg)$(a+b)$

We write the simple expression $a-b$, and consider it obviously the same as any of these:

$$a+(-b)=a-(+b)=a+(-1)\cdot b$$

but each is, in fact, a different expression with a different meaning. Given an ordinary positive number a, $-a$ is naturally defined as the result of subtracting a from 0. For a minute, to fix ideas, *don't* write $-a$, but use the notation (neg)a for $0-a$. Then note how complicated it is to define $a+b$ for all signs of a and b. Starting with a and b positive, Table II gives the sums and differences of a and (neg)a with b and (neg)b,

Understanding this table for the case of addition seems to be the first step in understanding and formalizing negatives. The second step is to extend subtraction to negatives so as to get the last column. This is contained in the rule:

$$a-(-b)=a+b, \text{ for all positive numbers } a, b.$$

The basic reason for this is that we want the identity $a-x+x=a$ to hold for all x, positive or negative or, in other words, subtraction should always cancel out addition. If we take x equal to $-b$, then replacing $a-(-b)$ by $a+b$ makes this identity hold. The argument one finds in some historical writings may be paraphrased as "taking away a debt of size x is the same as acquiring a new asset of size x", a fact obvious to any merchant. In any case, understanding of negatives up to this point seems to be a natural stage that one encounters in various historical documents. In modern terminology, while acknowledging that our modern words distort historical truth, one would paraphrase this stage by saying that it incorporates the idea that the integers, positive and negative are an abelian group under addition.

But multiplication of negatives is a subtler operation, the third and final step in the arithmetic of negatives. Modern notation again obscures the subtlety. When you write the simple identity $-a=(-1)\cdot a$, you are making a big step. Perhaps this is a

contemporary mathematician splitting hairs because historically this seems to have been assumed as completely natural by nearly every mathematician once they knew the rules for subtracting negative numbers (with the exception perhaps of Cardano and Harriot, see below). One difficulty in arguing for this rule is that there are not many simple cases of quantities in the world where the units of the two multiplicands allow us to infer the multiplication rule using our physical intuition about the world. Here are a number of ways of arguing that the identity $(-1)\cdot(-1) = +1$ must hold.

Method I: Use the basic, intuitively obvious, identity:

$$distance = velocity \times time$$

and argue that if you substitute:

(a) *velocity* = movement of one meter *backwards* per second, a negative number,

(b) *time* = second in the *past*, also negative,

(c) then one second ago, you were 1 meter ahead, i.e. *distance* $= +1$ meter.

This 'proves' $(-1)\cdot(-1) = +1$.

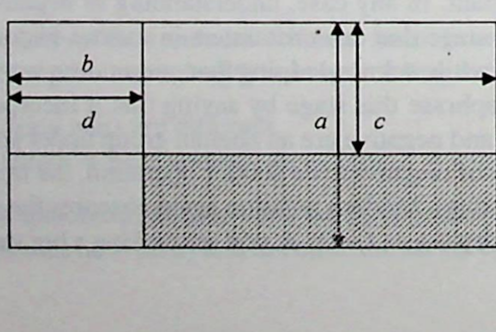
Method I′: I know of only one other real world situation where the rule is intuitively obvious. This variant of the previous argument concerns money and time. We use the simple equation obvious to any merchant describing the linear growth of a business's assets:

$$assets\ at\ time\ t = (rate\ of\ change\ of\ assets) \times (elapsed\ time\ t) + (assets\ at\ present)$$

Now suppose a business is *losing* \$10,000 a year and is going bankrupt right now. How much money did it have a year ago? Substitute $t = -1$, $rate = -10000$, $present$ $assets = 0$ and the obvious fact that *assets a year ago* $= +10000$ to conclude that $(-1)\cdot(-10000) = +10000$.

Method II: (as in Euclid's geometric algebra)

In Euclid, multiplication occurs typically when the area of a rectangle is the product of the lengths of its two sides. Consider the diagram below:



The big rectangle has area $a \cdot b$ but the shaded rectangle has area $(a-c)\cdot(b-d)$. Since the area of the shaded rectangle equals the area of the big rectangle minus the areas of the top rectangle and the left rectangle *plus* the area of the small top-left rectangle (which has been subtracted twice), we get the identity

$$(a-c) \times (b-d = ab - bc - ad + cd, \text{ if } a, b, c, d > 0, a > c, b > d$$

Now we use the idea that identities should always be extended to more general situations so long as no contradiction arises. If we extend this principle to arbitrary a, b, c, d, (which will bring in negative lengths and areas), we get for $a = b = 0$:

$$(-c)(-d) = +cd$$

This approach is probably the most common way to derive the multiplication rule. It can be phrased purely algebraically if you extend the distributive law to all numbers and argue like this (using also $0 \cdot x = 0$ and $1 \cdot x = x$):

$$1 = 1+(-1)\cdot 0 = 1+(-1)\cdot(1+(-1)) = 1+(-1)\cdot 1+(-1)\cdot(-1) = (-1)\cdot(-1).$$

<u>Method III:</u> Start with the multiplication

$$(\text{positive integer } n) \times (\text{any quantity } a) = (\text{more of this quantity } na)$$

(e.g. 4 × (quart of milk) = a gallon of milk), then by *subdividing* quantities as well as replicating them, you can define multiplication

$$(\text{positive rational}) \times (\text{quantity } a)$$

and by continuity (as in Eudoxus), define

$$(\text{positive real}) \times (\text{quantity } a)$$

What we are doing is interpreting multiplication of any quantity by a positive dimensionless real number as *scaling* it, making bigger or smaller as the case may be. Now if the quantity involved is signed you find it very natural to interpret reversing its sign as scaling by -1, i.e. to make the further definition:

$$(-1) \times (\text{quantity } a) = (\text{quantity } -a)$$

Now you have multiplication by any real number, positive or negative. In other words, the negative version of scaling is taking quantities to their opposites.

The core of this argument is the algebraic fact that the *endomorphisms of an abelian group form a ring* and we are constructing multiplication out of addition as *composition of endomorphisms*. This makes the third approach arguably the most natural to a contemporary mathematician trained in the Bourbaki style.

2. Negatives in Chinese and Indian Mathematics

We will discuss China first. The classic of Chinese mathematics is the *Jiuzhang Suanshu* (*Nine Chapters on the Mathematical Art*). Like Euclid, this is a compendium of the mathematical concepts and techniques which had been developed slowly from perhaps the Zhou (or Chou) dynasty (begins c.1000 BCE) through the Western Han dynasty (ending 9 CE). Unlike Euclid, it is a list of practical real world problems and algorithms for their solution, without any indication of proofs. Since then, the *Nine Chapters* had a long history of ups and downs, sometimes being required in civil service exams and sometimes being burned and nearly lost. Each time it was republished though, new commentaries were added, starting with those of the great mathematician Liu Hui in 263 CE and continuing through those in the English translation by Shen, Crossley and Lun[5]. Page numbers in our quotes are from this last edition.

Starting some time in the first millennium BCE, arithmetic in China began to be carried out using counting rods, which were arranged in rows using a decimal place notation. When doing calculations, different numbers were laid out by rods in a series of rows, forming a grid: a Japanese illustration of how they were used is shown in the figure below.



Figure 2. A Japanese illustration of calculation with counting rods

The section of the *Nine Chapters* in which negative numbers are introduced and used extensively is Chapter 8, *Rectangular Arrays*. This Chapter deals with the solutions of systems of linear equations and expounds what is, to all intents and purposes, the method of Gaussian Elimination. In fact, it is indistinguishable from the modern form. The coefficients are written out in a rectangular array of rod numerals and one adds and subtracts multiples of one equation from another equation until the system has triangular form. Examples as large as five equations in

[5] *The Nine Chapters on the Mathematical Art: Companion and Commentary*, Shen Kangshen, John N. Crossley, and Anthony W. -C. Lun, Oxford University Press, 1999.

five unknowns are worked. Naturally negative numbers appear all the time in such an algorithm.

As described in Liu's commentary, red rods or upright rods were used for positive numbers which he calls gains (*zheng*) and black rods or slanting rods for negative numbers which he calls losses (*fu*). He says "*red and black counting rods are used to cancel each other*". Curiously, his colors are the exact opposite of our Western accounting convention! Here is Problem 8 from this Chapter, p.409 in the Shen, Crossley and Lun edition:

> *Now sell 2 cows and 5 sheep, to buy 13 pigs. Surplus: 1000 cash. Sell 3 cows and 3 pigs to buy 9 sheep. There is exactly enough cash. Sell 6 sheep and 8 pigs. Then buy 5 cows. There is 600 coins deficit. Tell: what is the price of a cow, a sheep and a pig respectively?*

This means the three equations (all of which have negative coefficients as well as positive):

$$\begin{aligned} 2C + 5S - 13P &= 1000 \\ 3C - 9S + 3P &= 0 \\ -5C + 6S + 8P &= -600 \end{aligned}$$

The solution is found to be $C = 1200$, $S = 500$, $P = 300$. The *Nine Chapters* goes on rather mysteriously (p.404):

> *Method: Using rectangular arrays lay down counting rods for each entry to be added.*
>
> *The Sign Rule*
>
> *Like signs subtract; opposite signs add; positive without extra, make negative; negative without extra, make positive.*
> *Opposite signs subtract; same signs add; positive without extra, make positive; negative without extra, make negative.*

Liu's commentary explains: the first set of sign rules refers to subtraction of array entries, the second to addition. He goes on to clarify the meaning of the cryptic Sign Rule. In fact, the rule is precisely what we wrote out in Table II above for both addition and subtraction. What is clear is that negative numbers were analyzed and treated correctly as soon as the need arose, presumably for the first time anywhere in the world.

I cannot find in the Shen et al edition of the *Nine Chapters* any treatment of multiplication of negative numbers, although Martzloff[6] quotes the Chinese edition of Qian Baocong as saying: "*Rods of the same name multiplied by each other make positive. Rods of different names multiplied by each other make negative*". In any

---

[6] Jean-Claude Martzloff, *A History of Chinese Mathematics*, 2nd edition, Springer, 1997, page 203.

case, Liu's commentary, written in the 3rd century CE, makes the remark (p.405): "*Interchanging the red and black rods in any column is immaterial. So one can make the first entries of opposite sign.*" This is the correct rule for multiplication by -1.

Chinese algebra had a renaissance in the Song and Yuan (Mongol) dynasties. In particular, Zhu Shijie (c.1260–c.1320) extended the ideas of Gaussian elimination to the simultaneous solution of *polynomial* equations, inventing the equivalent of the resultant and using ever larger and more complex arrays of coefficients. At this stage, as one would expect, the full rules for negative arithmetic emerge quite explicitly as well those for the algebra of polynomials. Having a theory of negatives is the clear prerequisite for going further in the study of algebra. Zhu's algebra reached a stage not attained in Europe until the late 19th century.

I want to turn to India next. In every culture, one of the main reasons for the development of arithmetic – arguably the principle driving force – is the need of merchants to keep accounts. In fact, it is even hypothesized that arithmetic and writing itself emerged in the 3rd millennium BCE in Mesopotamia as a development of a crude system of tracking transactions of agricultural goods by means of small specially shaped and inscribed tokens[7]. By around the year 2000 BCE, one finds tablets from Ur[8] with a yearly summary accounting, showing budgeted and actual inputs (with value converted into a common unit of barley), budgeted and actual outputs, budgeted and actual labor and differences, shortfalls or profits! In India, very sophisticated principles of accounting were codified in Kautilya's comprehensive manual of statecraft, the Arthaśātra written in the 4th century BCE. The Arthaśātra covers in amazing detail every aspect of setting up and managing of a kingdom (including managing a special forest for elephants). In Book II, Chapter 6 and also in many later Chapters of Book II, Kautilya details how accounts are to be kept[9]. He describes a complete system of book keeping: he has a ledger for *income* with dates, times, payers, categories, etc. and a ledger for *expenditures* and finally a third ledger for *balances*. There are sections on auditing, insurance against theft, debtors, borrowings, mortgages, auditing, etc. and subtler accounting issues such as current vs. deferred receipts, how to account for price changes of items in inventory, fixed vs. variable costs. Although he does not use negative numbers explicitly, he is

[7]The pioneer here has been Denise Schmandt-Besserat, who has brought her life's work together in the multi-volume book *Before Writing*, volume I being *From Counting to Cuneiform*, University of Texas Press, 1992. In particular, she has "deciphered" the mysterious tokens found throughout the Middle East from roughly 8000 BCE to 3000 BCE, finding a simple method of accounting which merges seamlessly with highly developed cuneiform accounts in the 3rd millennium.

[8]See Chapter 5 in Richard Mattessich, *The Beginnings of Accounting and Accounting Thought: Accounting Practice in the Middle East (8000 B.C to 2000 B.C.) and Accounting Thought in India (300 BCE to the Middle Ages)*, Garland Publishing, 2000.

[9]See Chapter 6 in Mattessich, *Op. Cit.*, which is based on the book *Modern Accounting Concepts in Kautilya's Arthasastra* by Anjan Bhattacharyya, Firma KLM, Calcutta, 1988.

clearly aware of how accounts must sometimes show a deficit and that people may carry a negative net worth.

Although the Arthaśātra does not mention negative numbers explicitly, they appear full blown in Brahmagupta's treatise *Brâhma-sphuta-siddhânta* (628 CE). The development of mathematics in India in the first millennium CE is connected much more strongly to astronomy than to accounting. For much of this period, treatises covering both mathematics (*ganita*) and astronomy (the motion of the sun, moon and planets and their positions at a given time and place in the sky) and called *siddhantas* were composed. Many of these were in verse, highly compressed and cryptic, meant to be memorized and handed down generation by generation from teacher to student.

The *Brâhma-sphuta-siddhânta* includes two Chapters on mathematics which are a compendia of the mathematical concepts and techniques which had been developed over previous centuries. Here we find all the correct rules for arithmetic with negative numbers and in it *positive numbers are referred to as "fortunes", negative numbers as "debts"*. It appears that accounting led naturally to an arithmetic in which negative numbers took their natural place. Here are some quotations, showing first the rules we laid out in table I and then, significantly, going on to describe how to multiply negative numbers[10]:

> *[The sum] of two positives is positive, of two negatives, negative; of a positive and a negative [the sum] is their difference; if they are equal, it is zero. The sum of a negative and zero is negative, of a positive and zero positive, of two zeros, zero.*
>
> *[If] a smaller [positive] is to be subtracted from a larger positive, [the result] is positive; [if] a smaller negative from a larger negative, [the result] is negative; [if] a larger from a smaller, their difference is reversed – negative becomes positive and positive negative.*
>
> . . . .
>
> *The product of a negative and a positive is negative, of two negatives positive, and of positives positive; the product of zero and a negative, of zero and a positive, or of two zeros is zero.*
>
> *A positive divided by a positive and negative divided by a negative is positive; a zero divided by a zero is zero; a positive divided by a negative is negative; a negative divided by a positive is negative.*
>
> *Chapter 18, verses 30–34*

The only oddity seems to be his confident assertion that $0/0 = 0$. The rest is as clear and modern as one could wish for. It would be wonderful to know what considerations led Indian mathematicians in the late centuries BCE or the early centuries CE to these conclusions – especially for the multiplication of negative numbers. The predominately oral transmission of knowledge in the Vedic

[10]We quote from the translation by Kim Plofker in her book, *Mathematics in India, 500 BCE – 1800 CE*, Chapter 5, p.151.

tradition – and perhaps the difficulty of preserving perishable writing materials through yearly monsoons – has not left us with any record of these discoveries. They just appear full blown in Brahmagupta's summary. R.Mattessich has developed at length the idea that it was the highly developed tradition of accounting which led to the full understanding of negative numbers[11] but unfortunately no evidence for this plausible conjecture exists.

As in China, having negative numbers opened the way to deeper studies of algebra itself. Perhaps the deepest of these was the Indian work on Pell's equation $x^2 - Ny^2 = m$, especially finding solutions for $m = 1$. Brahmagupta himself made the first huge step, discovering the multiplication law arising from the factorization

$$x^2 - Ny^2 = \left(x + \sqrt{N}y\right)\left(x - \sqrt{N}y\right).$$

More exactly, he showed how from solutions of the equation for m_1, m_2, one gets one for their product $m = m_1 \cdot m_2$. Some centuries later, Jayadeva found a complete algorithm for constructing solutions with $m = 1$.

We find reflections of the Indian use of negatives in their astronomy too. As stated, the main goal of these scholars was not to develop mathematics for its own sake but to apply mathematics to predict the positions of the sun, moon and planets. An epicyclic theory is used and, for the planets, both a 'slow' and 'fast' correction is added to the mean motion of the planet (in our terms, one is due to the ellipticity of their orbit, the other to the shift from a heliocentric to a geocentric description). David Pingree[12] has hypothesized that through the intermediary of the Indo-Greek empire, some version of the pre-Ptolemaic Hipparchan theory of planetary motion reached India. What is quite striking is that in making these corrections the sine function *in all four quadrants* is understood. Hipparchus had computed tables of *chords*, which are fundamentally unsigned positive quantities. The Indian tradition shifts to sines (actually 'Rsines', sines multiplied a large radius and rounded to the nearest integer) and then it is natural to extend them from the first quadrant to the full circle. Here is a quote from the *Brâhma-sphuta-siddhânta*, Chapter 2, verse 16 describing the corrections made by adding or subtracting appropriate sine function corrections to the mean position:

> *(In successive quadrants) (in the slow case) negative, positive, positive, negative correction, otherwise in the fast case. (The sum) of two positives (is) positive, of two negatives (is) negative; of positive and negative (is their) difference, of equals (positive and negative is) zero.*[13]

---

[11] See Chapter 7 in Mattessich, *Op. Cit*.

[12] David Pingree, The History of Mathematical Astronomy in India, in *Dictionary of Scientific Biography*, Charles Gillespie editor, Scribner, 1978, pp.533–633.

[13] Translation by J. Shah (personal communication).

It would be nice if they had drawn a graph of the correction in all quadrants, i.e. of the sine function, to clarify this verse, but that was clearly not their *modus operandi*. But further evidence that the sine function was seen as being extended to more than one quadrant comes from the rational approximation of the sine in the first two quadrants given by Bhaskara I (7th century CE)[14]:

$$\sin(\theta) \approx \frac{16 \cdot \theta(\pi - \theta)}{5\pi^2 - 4\theta(\pi - \theta)}, \; 0 \leq \theta \leq \pi$$

This is an extraordinarily accurate approximation which would be hard to come up with if they had not grouped the first and second quadrant together.

Another natural place for using negative numbers is for coordinates, e.g. to measure the celestial latitude (perpendicular to the ecliptic), or the declination (perpendicular to the celestial equator), of a planet or star. Tradition, however, sanctifies describing latitudes and declinations as north/south instead of positive/negative and this is hard to change. But this latitude must often be put into formulas when converting from celestial coordinates to horizon based coordinates, e.g. when calculating the very important rising times of planets. At this point, rules for negatives again must be used. Here is an example from Brahmagupta's *Khandakhâdyaka,* Ch.6, verse 5[15]

> *Multiply the north celestial latitude by the equinoctial shadow and divide by 12; apply the quotient taken as minutes negatively or positively to (the longitude measured from) the orient and occident ecliptic points. When the celestial latitude is south, apply the resulting minutes to the same points positively or negatively.*

In modern terms (see Figure 3), he is computing $(longitude\ KA) \pm (latitude\ KV) \times \tan(\phi)$, where ϕ is the observer's latitude and distinguishing the cases where longitude is measured eastwards or westwards and where the planet's latitude is north or south.

An explicit interpretation of negatives as coordinates on a number line occurs later in the work of the 12th century Bhaskara II (so-called to distinguish him from the earlier 7th century Bhaskara I). He wrote an immensely popular textbook on Algebra, the *Lîlâvatî*[16]. The title was said by a Persian translator to be the name of Bhaskara's daughter and, although this is not made explicit in the book, it is full of verses addressed to the "beautiful one", "the fawn-eyed one", etc. Present day texts are so drab in comparison!

The remarkable passage is in verse 166 and again it is given without any fanfare stating that a new interpretation of negative numbers is being given. But, to my

---

[14]Bhaskara I, *Mahabhaskariya*, Ch. 7, verses 17–19.

[15]*The Khaṇḍakhâdyaka of Brahmagupta*, with the commentary of Bhaṭṭotpala, edited and translated by Bina Chatterjee, World Press, Calcutta, 1970, p.122–3.

[16]We follow the classic translation by H. T. Colebrooke, first published by in 1817 and subsequently reprinted in numerous editions.

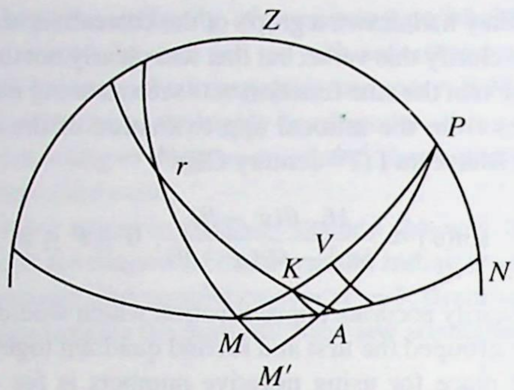


Figure 3. Diagram for calculation of rising time. Planet at V, rKA ecliptic, P north celestial pole.

knowledge, it is the first occurrence of the "number line", of using positive and negative numbers as coordinates on either side of an origin. Bhaskara is in the middle of a discussion of triangles and, specifically, given the three sides a, b, c of the triangle with a distinguished side c, the base, how to find the altitude and the position of the foot of the perpendicular dropped on the base. If you let x be the distance from one endpoint of the base to the foot, then $(c - x)$ is the distance from the other endpoint to the foot and Pythagoras's theorem tells us:

$$a^2 - x^2 = \text{altitude}^2 = b^2 - (c - x)^2$$

which gives us:

$$x = (a^2 + c^2 - b^2)/2c$$

In verse 166, he poses the problem:

> *In a triangle, wherein the sides measure ten and seventeen and the base nine, tell me promptly, expert mathematician, the segments, perpendicular and area.*

and his formula gives him $x = -6, c - x = 15$ (see Figure 4). Aha: what to do? Well, if you draw this triangle, you find the foot of the perpendicular lies outside the base. So what does Bhaskara say?

> *(The result 6) is negative, that is to say, in the contrary direction. Thus the two segments are found 6 and 15. From which, both ways too, the perpendicular comes out 8.*

This is stated so casually, as if it were common wisdom, that one can only conclude that this way of thinking about negative distances was well-known in his

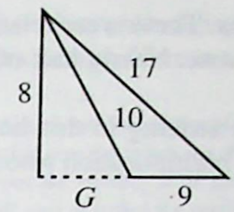


Figure 4. A triangle with a perpendicular falling outside the base, Bhaskara II

time. Nonetheless, as we will see, it doesn't occur in Europe before the work of Wallis near the end of 17^{th} century.

3. The Shunning of Negative Numbers, From Al-Khwarizmi to Galileo

I now turn to the Arab and Western treatment of negative numbers. To keep the story within bounds, I will pick a small selection from the many figures who might be discussed, those who seem to me key figures in the story or who exemplify a particular stand.

Al-Khwarizmi (c.790–c.840)

It is repeated everywhere that the Indians invented zero and place notation and that the Arabs learned it from them and later transmitted this to Europe. It's bizarre that such a misunderstanding should be widespread but in fact, the Babylonians invented place notation (albeit using base 60) and their arithmetic was used by many Greeks, e.g. Ptolemy. I hope I have made the case that the most substantial arithmetic discovery of the Indians – and independently the Chinese – was not merely that of zero but the discovery of negative numbers. Sadly this discovery was not absorbed in any but a superficial way by the Arabs.

Al-Khwarizmi (whose full name was Abu Ja'far Mohammad ibn Musa Al-Khwarizmi) was familiar with Indian mathematics and astronomy and apparently with Brahmagupta's *Brâhma-sphuta-siddhânta* written some 200 years earlier. He worked under the patronage of the caliph Al-Mamun about whom he says "*That fondness for science, . . . , that affability and condescension which he* (the caliph) *shows the learned . . . has encouraged me to compose a short work on calculating by Completion and Reduction .. such as men constantly require in cases of inheritance, legacies, partition, law-suits and trade . . .* "[17] His book on Algebra is entitled *Al-jabr w'al muqabala* which refers to the operations of completion and reduction with

[17] *The Algebra of Mohammed ben Musa,* Translated by Frederic Rosen. Facsimile reprint of 1831 edition by the Oriental Translation Fund, London, Adamant Media Corporation 2002.

which he simplified his equations. These were relations between an unknown, its square and constants, given in prose. Nearly half of his book concerns incredibly complex inheritance cases.

I find three things especially striking in this book. Firstly, negative numbers appear only once, in a section on multiplication whose goal appears to be to explain the identity

$$(a - c) \cdot (b - d) = ab - ad - bc + cd$$

and justify it by geometry, just as in our discussion of "Method II" for multiplying negative numbers. But then they are never mentioned again. The second striking thing is that quadratic equations always have positive coefficients and thus belong to three types:

1. $ax^2 + bx = c, a, b, c > 0$ (referred to as "roots and squares are equal to numbers")
2. $ax^2 + c = bx, a, b, c > 0$ ("squares and numbers are equal to roots")
3. $ax^2 = bx + c, a, b, c > 0$ ("roots and numbers are equal to squares")

This separation of cases continues down through the whole European tradition through De Morgan. An equation, in short, must be an identity between two positive numbers. Thirdly, he discusses exactly the same problem that Bhaskara II was to take up: finding altitudes of triangles whose sides are given. But, unlike Bhaskara, all the examples he treats have the foot of the perpendicular *inside* the base so this big clue about negatives never comes up.

Leonardo of Pisa (1170–1250)

Leonardo of Pisa was one of the first Europeans to master the Arab arts of calculation, including the use of Indian symbols and place notation. He wrote a remarkable book, his *Liber Abaci* (Book of Calculation), in which the rules for all the basic arithmetic operations are laid out in great detail and exhaustively illustrated by numerical examples. This occupies the first half of his book which is essentially what we would call a primer. But he deals exclusively with the arithmetic of positive integers and positive fractions. His section on subtraction is entitled *On the Subtraction of Lesser Numbers from Greater Numbers*.

As in the Indian tradition, accounting was one of the principle stimuli for the development of arithmetic in the Middle Ages and much of the book deals with the arithmetic of money, goods and possessions. The second half of the book treats a huge number of "word problems" involving goods and money. He is following a curious tradition going back to Diophantus (and found in Chinese and Indian works also) of what, to modern eyes, are quite bizarre artificial "word problems"

involving a group of people who, after exchanging various sums of money, have sums satisfying some linear relationships. Here is an example[18]:

> *Three men had pounds of sterling, I know not how many, of which one half was the first's, one third was the second's and one sixth's was the thirds; as they wished to have it in a place of security, every one of them took from the sterling some amount, and of the amount that the first took he put in common one half, and of it that the second took, he put in common a third part, and of that which the third took, he put in common a sixth part, and from that which they put in common every one received a third part, and thus each had his portion.*

In modern algebra terms, if S is the sum of sterling and x_1, x_2, x_3 are the sums which the three men took, so that $(x_1/2 + x_2/3 + x_3/6)$ is what "*they put in common*", then the last sentence "*each had his portion*", sets up three equations:

$$\frac{x_1}{2} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{S}{2}$$
$$\frac{2x_2}{3} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{S}{3}$$
$$\frac{5x_3}{6} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{S}{6}$$

This is only one of hundreds of such problems. He develops methods of laying out the coefficients in rows and manipulating the numbers to get the answer. In the above, the 'answer', is the smallest set of relatively prime x's which solve these three homogeneous equations in 3 unknowns. Leonardo has a rather awkward and special version of the Chinese algorithm for solving linear equations in many unknowns.

Now most of his problems are set up so all the numbers which occur are positive. But not all! First of all, negative numbers can arise in the course of the calculation. He then says things like[19]:

> [he is in the middle of an algorithm]. . . *and from the* 240 *you subtract* 288 *leaving minus* 48, *and this I say because the* 288 *cannot be subtracted from the* 240; *from this* 48 *you take* 1/3 *for the* 1/3 *of the second position; there will be minus* 16 . . . .

He is getting close to the red and black rods of the Chinese, but these examples are few and far between and are not pursued very far. In a few other cases, the answer itself is negative. For example, after solving the problem described in the first quote, he varies the proportions of S owned by the three men to 1/2, 2/5 and 1/10. In this case, the solution is $x_1 = 326, x_2 = 174$ and $x_3 = -30$. The setting of the problem, that all the x's are amounts of money, comes to his rescue. The third man, he says,

---

[18] Leonardo of Pisa, *Liber Abaci*, p.415 of the English translation by L. Sigler, *Fibonaccis' Liber Abaci*, Springer-Verlag, 2002.

[19] Ibid, p.419.

does not take anything from the sum S which they share but instead puts in an additional 30 pounds of his own "proper" money: there were 470 pounds in all, and when they "wanted to have it in a place of security", the third man *added* 30, the first man took 326 and the second took 174. When money is concerned, negative quantities can always be given a simple meaning!

Leonardo is making the first tentative steps towards enlarging the number system to include negatives. With money, he is comfortable with assets and debts, giving and taking. But his examples are few and he never makes explicit rules for extending arithmetic.

Nicole Oresme (1323–1382)

Nicole Oresme was a mathematically inspired scholastic, working in Paris in the mid-14th century. He made a giant stride taking geometry beyond Euclid. In his great book, *Tractatus de configurationibus qualitatum et motuum* (*Treatise on the configurations of qualities and motions*)[20], he proposed considering all intensities which varied in time and whose values at different times could be compared by a proportion. To any such quality, he proposed constructing a *graph*. First he took a line segment, called the *subject,* whose points represented the interval of time over which the quality was varying. This, in itself, was a radical departure from Euclid: now space was being used *analogically*, as a substitute for time. Then he proposes erecting line segments perpendicular to the subject whose lengths had the same proportions as the qualities being graphed:

> *Therefore, every intensity which can be acquired successively ought to be imagined by a straight line perpendicularly erected on some point of the space or subject of the intensible thing, e.g. a quality. For whatever ratio is found to exist between intensity and intensity of the same kind, a similar ratio is found to exist between line and line, and vice versa. . . . Therefore, the measure of intensities can be fittingly imagined as the measure of lines.* (Oresme, I.i)

He talks about graphing many things (although he never gathers data or actually goes beyond making simple cartoons of his graphs – see Figure 5). In particular, he discusses graphing velocity, temperature, pain and grace (of a soul). Some of these are clearly positive quantities by nature, e.g. pain and grace. He is interested in contrasting intensities which are constant (graph (a) in figure), intensities which vary at a constant rate (graph (b) in the figure) and intensities which are more complex (graphs (c) in figure). For example the grace of a soul '*occupied by many thoughts and affected by many passions*' will be difformly difform – his name for type (c). On the other hand, velocities can clearly change sign and, as for temperature, he even considers there to be complementary intensities of hotness and coldness. For

[20] Translations are from Marshall Clagett's translation, *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, University of Wisconsin, 1968.

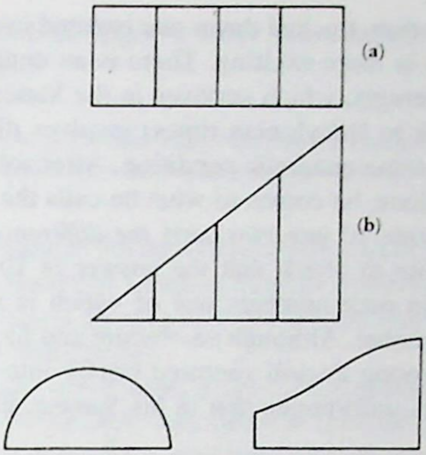


Figure 5. Oresme's examples of graphs

temperature, hotness might have a graph with values $f(x)$ and coldness a graph with values $C - f(x)$. In other words, he adds a suitable positive constant so as to make every intensity positive everywhere.

Because his graph is the whole area, not simply the curve at the tips of the his line segments, he cannot have a graph which goes from positive to negative, crossing the 'subject'. This is especially striking because at one point he makes a catalog of various types of difformly difform graphs: but no graph in the catalog is, for example, regularly oscillating like a sine wave. He even hints at the fact that the area of the graph of velocity is the distance traveled, the fundamental theorem of calculus, but to make his picture, the velocity cannot change sign: no backtracking. Oresme has gone beyond Euclid in a striking way but he cannot make the further leap of allowing negative values for an intensity.

Luca Pacioli (1445–1517)

Pacioli's importance is not due to his discoveries but to the fact that he wrote an encyclopedic work *Summa de arithmetica, geometria, proportioni et proportionalita* which summarizes the contemporary knowledge of arithmetic, geometry and especially accounting. The work's greatest influence was due to its description of double-entry book keeping which was a key step in the expansion of the international business enterprises which characterized the Renaissance. Here we find a small number of linear equations involving amounts of money whose solution is negative. As in Leonardo, when the result was a negative number, it is described as a debt. In one case, the price of an egg comes out negative – owning the egg puts you in debt so the sellers are paying you to take their eggs.

Sesiano (*op.cit.*), however, tracked down one isolated instance of a problem in Pacioli's writings which is more exciting. There is an untitled manuscript, written for his students in Perugia, which survives in the Vatican[21]. A standard class of problems (going back to Babylonian times) involves dividing a number into two parts which satisfy some quadratic condition. After solving some such problems with positive solutions, he comes to what he calls the *bellissimo caso.* This example asks you to divide 10 into two parts *the difference of whose squares* is 200. The reader may like to check that the answer is $10 = 15 + (-5)$. Here is a problem not only in pure numbers one of which is negative but requiring squaring this negative number. Although an obscure and forgotten footnote to history, it seems that the young Pacioli ventured briefly into uncharted territory in a truly original way. It is unfortunate that in his *Summa*, he did not pursue these ideas.

Girolamo Cardano (1501–1576)

The only reason to include Cardano is that he wrote the book *Ars Magna*[22], so we can analyze how he thought, how he looked on negative as well as imaginary numbers. The solution of cubic equations was due to Scipione del Ferro, Professor of Mathematics at Bologna around 1515, and the solution of the quartic to Cardano's student Ludovico Ferrari. Cardano himself was an arrogant man, a compulsive gambler, who led a wild life of ups and downs. That he computed the odds of various sorts of gambling was arguably his greatest mathematical achievement.

If Al-Khwarizmi had spun out the solutions for quadratic equations in to many different cases, Cardano really went to town describing how to solve 13 distinct cases of cubic equation (and 44 types of derivative cases). Why so many? Because (a) the coefficients all had to be positive and (b) the equation had to equate a positive quantity to another positive quantity. The many sections are entitled things like "*On the cube and square equal to the first power and number, generally*". Nonetheless, he did recognize that some of his equations had negative solutions: these he called

> "*fictitious (for such we call that which is a debitum or negative)*"

but he does very little with such roots, ignoring them systematically. But in the later Chapter, "*On the rule for postulating a negative*", he does explore a bit what algebra

[21] Cod. Vat. lat. 3129.

[22] Quotations are from the 2007 Dover reprint *The Rules of Algebra: (Ars Magna)*, translated by T. Richard Witmer.

can do for you if you admit negative roots. His example of a problem requiring negative numbers is this:

> *The dowry of Francis' wife is 100 aurei more than Francis' own property, and the square (?) of the dowry is 400 more than the square of his property. Find the dowry and the property.*

This works out to give Francis –48 aurei of property, that is, he is in debt 48 aurei, but fortunately is getting a dowry of 52 aurei. Here he correctly identifies the negative solution with a debt. This is an excellent illustration although squaring a sum of money is a pretty weird thing to do.

There would little else to say except for the curve ball that was thrown to Cardano: for all cubic equations which have only one real root, del Ferro's formula worked like a charm. But if there were three real roots (the other possibility, known as the *casus irreducibilis*), it gave an apparently meaningless result. His formula for the roots of the equation $x^3 + ax + b = 0$ is:

$$x = \sqrt[3]{\left(b/2 + \sqrt{-D/4.27}\right)} + \sqrt[3]{\left(b/2 - \sqrt{-D/4.27}\right)}, \text{ where } D = -4a^3 - 27b^2$$

D, the discriminant, is equal to the square of the difference of all pairs of distinct roots, hence it is positive if all the roots are real. So we need to find the a square root of a negative number even though in the end we only want the real number x. Cardano struggled unsuccessfully with what this might possibly mean.

His one attempt to deal with these complex expressions is in the same Chapter, "*On the rule for postulating a negative*" mentioned above. Here he considers problems which have complex roots, such as the following:

> *Divide 10 into two parts the product of which is 40.*

The usual quadratic formula gives the two parts as $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. This is also the answer his math gives him and which he puts in writing in his book but he doesn't attribute much meaning to it. He makes his famous comment:

> *So progresses arithmetic subtlety, the end of which, as is said, is as refined as it is useless.*

At the end of this Chapter, he gives a third type of example where he reasons incorrectly with products of a real and an imaginary. In a later edition, he added an appendix *De aliza regula liber* in which he flirted with the idea that maybe $(-1)^2 = +1$ was wrong. Why not try $(-1)^2 = -1$? Between 'fictitious' negative numbers and useless imaginaries, you get the sense that Cardano was at sea.

Galileo (1564–1642)

Perhaps mathematicians were stuck thinking that negative numbers were fictitious but surely physicists who were actually measuring things in the real world, had a clearer view? Arguably, Galileo's great contribution to physics was his recognition that momentum was a key property of objects, that it was constant when no forces were acting and that the force of gravity acting on projectiles and falling bodies changed their momenta at a constant rate, not their positions. As an old man, when the Pope commuted his sentence for heresy to house arrest, he wrote down these theories in his *Dialog concerning Two New Sciences*[23]. He starts off with his foil Simplicio getting put down again and again by Galileo's mouthpiece Salviato. But by the Fourth Day, Galileo lapsed into a more standard Euclid-style exposition and puts out the centerpiece of his theory: the demonstration that a projectile follows a parabolic arc under the force of gravity. Here was something he had actually experimented with and he was on solid ground, theoretically as well as experimentally. Figure 6 is an excerpt from his notebooks working on projectiles.

The central assertion in these dialogs is that gravity endows the projectile with a constant downward acceleration. Thus its vertical velocity will be positive going up, zero at the peak and negative coming back down. It is a linear function changing from positive to negative. The math couldn't be simpler – *if you are willing to use negative numbers*.

What does Galileo do? His main result is:

Theorem 1. *A projectile which is carried by a uniform horizontal motion compounded with a naturally accelerated vertical motion describes a path which is a semi-parabola.*

Note that he uses a semi-parabola: the half of the parabola in which height is a monotone function of time. Considerably later, after a long discussion of the time and distance of the semi-parabolic arc carrying the projectile to the ground, he reverses time without any discussion and concludes that the rising phase of a projectile is also a semi-parabola.

The discussion continues on optimal angles at which to fire guns. But the astonishing point is that he never talks about the whole parabolic arc, with ascending and descending halves and how there is constant downward acceleration throughout the path. *All* the diagrams in the book resemble the figure from his notes: a semi-parabola with some auxiliary chords and tangents. He analyzes the geometry of the semi-parabola and the physics of a falling body and then asserts without any discussion that one can reverse the direction of motion from a fall to a climb – nothing else. That the velocity changes at the apex from positive to negative is not stated anywhere.

---

[23]Quotations are from the 1956 Dover edition, *Dialogues Concerning Two New Sciences, translated by* by Alfonso De Salvia Henry Crew.

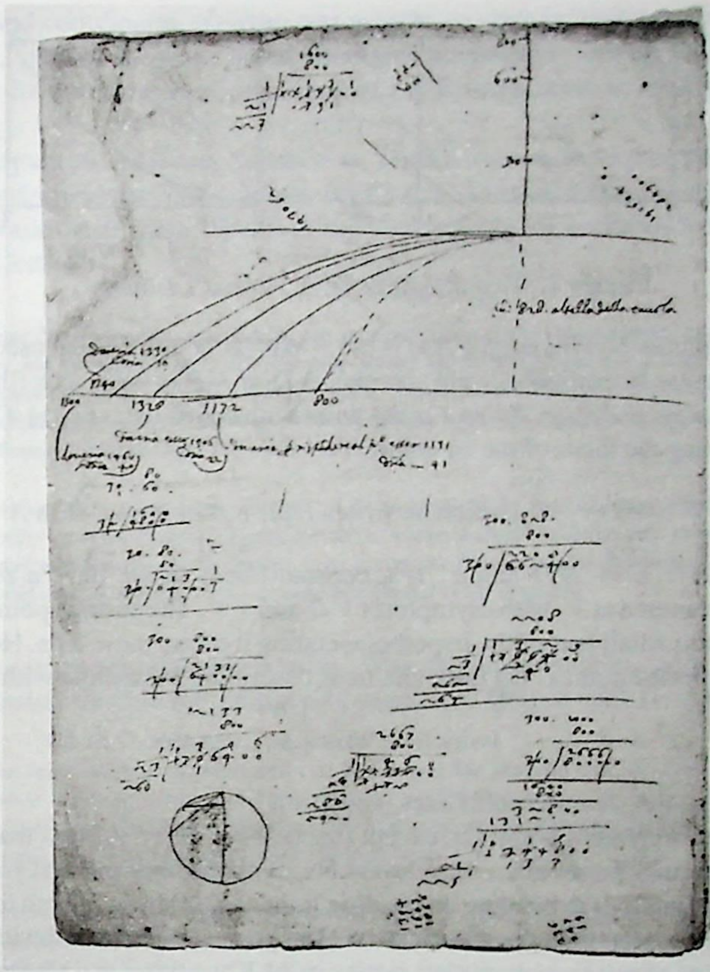


Figure 6. Galileo's notes on projectiles

Fermat (1601–1665)

Fermat and Descartes, at essentially the same time, had the idea of introducing coordinates into the plane and connecting geometric loci with polynomial equations in two variables. Plane curves are not confined to the positive quadrant, so one might expect that their logic would have pushed them to allow their variables to take on both positive and negative values. But no! Their coordinates were only in a positive quadrant and the other parts of a curve were treated separately if at all.

Below are two figures from Fermat's paper on the subject, *Ad Locos Planos et Solidos Isagoge*, (*Introduction to Plane and Solid Loci*). Incidentally, *plane* loci meant lines and circles, *solid* loci meant the other conic sections, terminology which dates from Greek times.

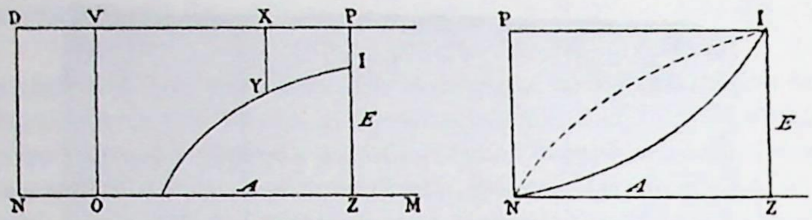


Figure 7. Two diagrams from Fermat's *Isagoge*

In these figures N is the origin, NM is the x-axis (although Fermat used the letter A, not x because his variables were vowels), ND or NP is the y-axis (the letter E for Fermat), $x = NZ$, $y = ZI$ so I is the point with coordinates (x, y). On the left, he is describing the locus of the equation:

$d^{*} + x \cdot y = r \cdot x + s \cdot y$ (which he writes Dpl. $+ A$ in E aeq. R in $A + S$ in E).

Here $s = NO, r = ND$ and d^{*} is a constant area, so we have a rectangular hyperbola, centered at V, with asymptotes VO and VP. The curious point is that he draws only this small part of the hyperbola, cutting it off on the x-axis. He also cuts it off at the plotted point I. On the right, he is describing a parabola with equation:

$x^2 = d \cdot y$ (which he writes Aq .aequatur D in E)

Again, he cuts the locus off at his axes (and at I).

Descartes' treatment is similar, except that he does say in the text that there are multiple orderings possible for the relevant points on the axes and that you must set up different equations depending on the directions and ordering of both the variable point and the constants in the construction. The goal is to make both sides of your equation sums of positive quantities, just as in Al-Khwarizmi and Cardano's work on quadratic and cubic equations. Note that this is how Fermat's version of the equation for the hyperbola reads.

4. Clarifying the Muddle: Wallis and Newton

So when did European mathematicians begin to make their peace with negative numbers? The first treatment which seems to me quite modern is that of John Wallis (1616-1703), Professor of Mathematics at Oxford. He published his *Treatise on Algebra*[24], written in English, in 1685. This was just two years before Newton published by his earth-shaking *Principia Mathematica* and well after Newton had done his major work in mathematics. In his mathematical notes, where he used

[24] Available online at http://eebo.chadwyck.com through subscribing universities.

algebra and coordinates, Newton was equally modern in his treatment of negative numbers, putting them on equal footing with positive numbers. So we should attribute the first clear European view of negative numbers to Wallis and Newton equally.

In Chapter 16, *Addition, Subduction, Multiplication and Extraction of Roots in Specious Arithmetic*, Wallis defines negative numbers as nicely, simply and clearly as you could wish (here '*Specious*' is Viete's term for arithmetic with variables given by letters):

> *To these Notes, Symbols or Species are prefixed (as occasion requires) not only numeral figures, but the signs + and – (or plus and minus), the former of which is a Note of Position, Affirmation or Addition; the other of Defect, Negation or Subduction: According as such Magnitude is supposed to be, or to be wanting. And where no such Sign is, it is presumed to be Affirmative and the sign + is understood.*
>
> *And accordingly these Signs are still to be interpreted as in a contrary signification. If + signify Upward, Forward, Gain, Increase, Above, Before, Addition, etc. then – is to be interpreted of Downward, Backward, Loss, Decrease, Below, Behind, Subduction, etc. And if + be understood of these, then – is to be interpreted of the contrary.*

In this quote, the capitalization is his. With this understanding of negatives, how does he justify the rule for multiplying negatives? Here is what he says:

> *For the true notion of Multiplication is this, to put the Multiplicand, or thing Multiplied (whatever it be) so often as are the Units in the Multiplier. ... and this, whatever the thing Multiplied, Positive or Negative: for there may well be a Double Deficit as a Double Magnitude; and −2A is as much the Double of –A as +2A is the Double of A. ...*
>
> *But in case the Multiplier be a Deficit or Negative quantity; suppose −1; then instead of Putting the Multiplicand so many times, it will signify so many times to Take away the Multiplicand. ... so that + by – makes –; But to Multiply –A by −2 is twice to take away a Defect or Negative. Now to take away a Defect is the same as to supply it; and twice to take away the Defect of A is the same as twice to add A or to put 2A ... : So that – by – (as well as + by +) makes +.*

As far as I know, this is the first place in Western literature in which the rule of signs is not merely stated but explained so clearly. After this, when he gets to writing out the formulae for roots of equations, he no longer has to separate all these cases which we saw in Al-Khwarizmi and Cardano. For the quadratic he writes:

$$\text{Given the equation, } x^2 \pm 2bx = \pm c^2$$
$$\text{the roots are } x \pm b = \sqrt{\pm c^2 + b^2}$$

(I have only changed his variable from a to x and noted squares by using e.g. c^2 for his cc.) Note that he follows Euclid is making all terms homogenous – so that, for

example, x, b, c can all be lengths and the equation relates an area to an area. For this reason, he needs the symbol $\pm$ in front of the c^2.

Finally, Wallis gives what I believe is the first explicit use of the full number line, positives to the right, negatives to the left, in Western literature:

> *Yet is it not that Supposition (of Negative Quantities) either Unuseful or Absurd when rightly understood. And though, as to the bare Algebraick Notation, it import a Quantity less than nothing: Yet, when it comes to a Physical Application, it denotes as Real a Quantity as if the Sign were +; but to be interpreted in a contrary sense.*
>
> *As for instance: Supposing a man to have advanced or moved forward (from A to B) 5 yards; and then to retreat (from B to C) 2 yards; If it be asked, how much had he Advanced (upon the whole march) when at C? I find . . . he has Advanced 3 Yards. But if, having Advanced 5 Yards to B, he thence retreat 8 Yards to D; and it then be asked, How much is he Advanced when at D, or how much Forwarder than when he was at A: I say –3 Yards. . . . That is to say, he is advanced 3 Yards less than nothing. . . . (Which) is but what we should say (in ordinary form of Speech), he is Retreated 3 Yards; or he wants 3 Yards of being so Forward as he was at A.*

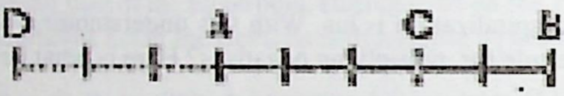


Figure 8. Wallis's illustration of the "number line"

Newton, as one would expect, had a full command of negative numbers and all their uses. He wrote lecture notes on arithmetic, algebra and geometry at some point, presumably early in his career. They were first published (without his approval) in 1707 and later translated into English with the title *Universal Arithmetick*. Here he introduces negative numbers at the very beginning with the following sentences[25]:

> *Quantities are either* Affirmative, *or greater than nothing; or* Negative, *or less than nothing. Thus in human affairs, possessions or stock may be called affirmative goods, and debts negative ones. And so in local motion, progression may be called affirmative motion, and regression negative motion; because the first augments, and the other diminishes the length of the way made. And after he same manner in geometry, if a line drawn in a certain way be reckoned for affirmative, then a line drawn the contrary way may be taken for negative.*

Later on, he discusses multiplication and is very clear that pure numbers arise as ratios of quantities with the same dimension and one can either multiply a quantity with a dimension by a pure number, getting another such quantity or multiply two

[25]Page 3 of the second edition published in 1728.

pure numbers. He states the rule for the sign of the product simply as "*. . . making the product* Affirmative *if both factors are Affirmative or both Negative; and* Negative *if otherwise.*" Unfortunately, he says nothing about why one should believe in this rule.

Whereas Fermat had given a systematic study of quadratic equations in two variables showing that they all defined conic sections and Descartes had introduced several cubic equations giving new curves (notably the "Cartesian parabola" and his "Folium"), Newton went on to look at all possible cubics, in an article entitled "Curves" in *Lexicon Technicum* by John Harris published in London in 1710. He classified them into 72 types and sketched them. *Without hesitation, he used all four quadrants of the plane* and plotted all roots (x, y), positive and negative. Here is an example:

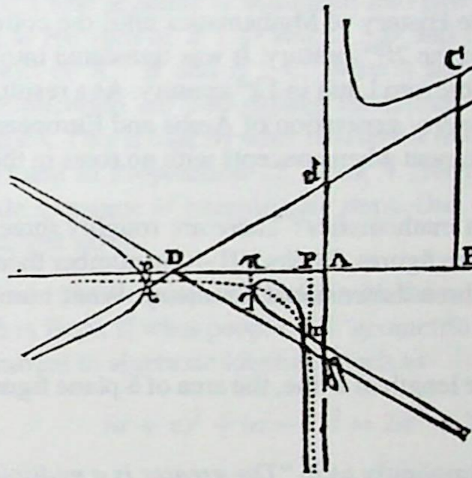


Figure 9. One of the 72 types of cubic curves plotted by Newton

After Wallis and Newton's work, a modern arithmetic with negative numbers was widely accepted in Continental Europe, where there was an explosion of mathematical research during the Enlightenment. In England, curiously, the resistance to negative numbers continued for some 150 years, culminating in De Morgan. A long debate ensued between those who accepted them and those that didn't, a story which is beautifully described in Pycior's book that we have cited. In the end, De Morgan and Hamilton founded the general theory of fields and negative reals took their place in the greater world of complex numbers and quaternions.

5. Two Factors in the World View of 15^{th}–17^{th} Century Europe

I hope I have proven my point that Europe in the 16^{th} and 17^{th} centuries resisted expanding their numbers to include negatives in a way which calls for some explanation. China and India both seem to have moved naturally to this bigger domain of

numbers when the occasion presented itself. I want to make the case that the European reticence was due to two factors. The first was the overwhelming importance of Euclid in defining what is and what is not mathematics and the fact that negative numbers had no place in Euclid's view of mathematics. The second is that, at the time negative numbers should have been accepted, imaginary numbers cropped up too and the idea arose that both negative and imaginary numbers had the same twilight existence. It was because of negatives that square roots had a problem, so maybe it was best to consider them both as second class citizens of the world of numbers.

Euclid's *Elements* were written in the newly founded school/library at Alexandria around 300 BCE and integrated the mathematical ideas of Theaetetus, Eudoxus and many others in a systematic treatise. It is written in a monolithic theorem/proof style not seen again in the History of Mathematics until the collective 'Bourbaki' composed their treatise in the 20th century. It was translated into Arabic in the 8th century CE and from Arabic into Latin in 12th century. As a result, it came to define what mathematics is for every generation of Arabs and Europeans, arguably until Newton and the Enlightenment when concepts with no roots in the Elements began to take center stage.

But what is Euclidean mathematics? There are roughly three parts to the *Elements*: Books I–VI on plane figures, Books VII–X on number theory and irrationals and Books XI–XIII on three dimensional geometry. What numbers occur in the *Elements*? Here's a list:

1. "*magnitudes*": the length of a line, the area of a plane figure and the volume of a solid figure
2. positive integers implicitly as in "*The greater is a multiple of the less when it is measured by the less*" (definition 2, Book V) and explicitly as in "*A number is a multitude composed of units*" (definition 2, Book VII). Note that the number is still a length but, because he always has a "unit" around when studying numbers, it becomes in effect dimensionless.
3. ratios as in "*A ratio is a sort of relation in respect of size between two magnitudes of the same kind*" (definition 3, Book V).

Note that none of these concepts give numbers which can be negative or even zero. What sort of arithmetic does Euclid have for these numbers?

1. Magnitudes are clearly added and subtracted (so long as the result remains positive), lengths are multiplied to give areas and volumes, etc. But "units" are only introduced in Book VII and there are no actual calculations and certainly no approximations (e.g. for π).
2. Positive integers are also added and subtracted and multiplication is defined in "*A number is said to multiply a number when that which is multiplied is*

added to itself as many times as there are units in the other and thus some number is produced".

3. Adding and multiplying ratios is the main goal in the extremely abstract Book V, which is said to be the work of Eudoxus. Book V begins with defining when two ratios are equal. For any ratio given by two lines A and B, he considers which multiples satisfy $nA > mB$ and which satisfy $nA < mB$. Of course, this is the 'cut' Dedekind re-introduced in the 19th century to *construct* real numbers from rationals. Here Eudoxus doesn't need to define real numbers – they are ratios given by geometry. What he needs to do is to define equality of ratios and he does this by requiring that their associated cuts are the same. Addition and multiplication of ratios are both implicit in that (a) if a line segment A is divided into two parts B and C then $A : D$ is going to the sum of $B : D$ and $C : D$ and (b) $A : C$ is to going to be the product of $A : B$ and $B : C$. What is not at all clear is that addition and multiplication are *well-defined* operations on the equivalence classes called ratios. This is exactly what is asserted in Proposition 24, Book V (for addition) and in Proposition 22, Book V (for multiplication) after a long and subtle sequence of intermediate steps. One stands amazed at Eudoxus' mathematical skills.

How about algebra, identities and formulas with the arithmetic operations? Euclid studies at length in Book II what people call 'geometric algebra', a series of propositions which amount to algebraic identities such as

$$(a + x)^2 + (a - x)^2 = 2a^2 + 2x^2$$

which is essentially the content of Proposition 9, Book II. Now what about the solutions of quadratic equations? This seems to be essentially what the lengthy and confusing Book X is all about. As Heath points out in his introduction to Book X, Euclid's classification of *binomials* and *apotomes* can be read as a systematic study of all the *positive* roots of all possible quadratic equations. This sets the stage for the separation of cases in treating roots of polynomial equations in all the works we have reviewed.

All in all, if you are going to start with Euclid, you are not going to be predisposed to introduce negative numbers in to your calculations. He has gone to extraordinary lengths to reduce arithmetic and algebra to geometry and thoroughly inoculate it against negatives. It is worth looking briefly at what else was known at 300 BCE which Euclid did *not* put in his book. There is apparently an unbroken tradition starting in Babylon in 1800 BCE and continuing through Ptolemy of calculating with the sexagesimal equivalent of decimals and approximating e.g. $\sqrt{2}$ and π to many sexagesimal places. Moreover, there was also a tradition also going this far back of solving quadratic equations by algorithms – described in words but exactly equivalent to the quadratic formula. Euclid, in other words, distanced himself from

a rich numerical tradition and consciously, it would seem, purified his version of mathematics.

The Europeans, then, had the benefit of this shining example of pure math and of the wonderful deductive logic on which they built. But it was hard to go beyond it in any radical way, to model other phenomena in the real world which cried out for negatives. Euclid was both the strength and the weakness of the European mathematical world of the 16th and 17th centuries.

But I think there is a second factor behind the slow acceptance of negatives which ought to be considered. As soon as one accepted -1, the algebra of the day thrust upon you formulae requiring its square root and this was truly inexplicable. The fate of -1 and i were inseparable. Cardano's book makes this very clear. We have already quoted from Chapter 37, near the end of his book, entitled *On the Rule for Postulating a Negative*. The Chapter starts with the sentence:

> *This rule is threefold, for one either assumes a negative, or seeks a negative square root, or seeks what is not.*

He is essentially equating three follies, all problematical. That he later entertained the idea that perhaps $(-1)^2$ ought to be equal to -1 shows how he viewed the problems as intertwined. Harriot (1560–1621) also played with both possibilities, as in the poem:

> *Yet lesse of lesse makes lesse or more,*
> *Use which is best keep both in store*
> ......

(Here '*lesse of lesse*' means multiplying -1 by -1 and he asks in line 1 whether this should equal -1 or $+1$).

Even if you didn't accept -1, the *casus irreducibilis* mentioned above, the case of cubic equations with three real roots, was a bone in the throat of algebraists. As long one of these roots was positive, you really ought to have a formula for this root. But the formula of del Ferro for solving cubics requires in this case that you take the square root of a negative number in an intermediate step. Of course, the imaginary parts of the resulting complex expressions will cancel at the end but not before. The full story of this problem is quite ironic. Viète in 1593[26] discovered that trisecting an angle was equivalent to solving the special cubic equation which belongs to the *casus irreducibilis*:

$$x^3 - 3x = b$$

[26] In his *Supplementum Geometriae*. A full treatment is in *Theoremata ad sectiones angulares* [Opera pp 287–304]

and he showed how to reduce the general *casus irreducibilis* to this special case. Thus he reduced a famous unsolved algebraic problem to a famous geometric one, unsolved in the sense that no ruler and compass construction was known (nor exists). At the same time, Bombelli proposed that Cardano's formula could make sense if you solved

$$\left(x + \sqrt{-y}\right)^3 = a + \sqrt{-b}$$

So trisecting an angle was related to taking complex cube roots – but no one put these together for a long time by finding the geometric meaning of complex numbers. Later we have Wallis, knowing the geometric meaning of negatives as the left half line, searching for a two dimensional geometric interpretation of imaginary numbers. There was a big clue on the table if anyone had linked Viète's trisection with taking cube roots of complex numbers. I believe it was Euler who finally worked out complex exponentials and made the link between these two. Oddly enough, even then Euler did not make explicit the geometric interpretation of complex numbers, leaving this to Wessel, Gauss and Argand.

Finally, there is also the issue of a psychological explanation for avoiding negative numbers. As Tversky and Kanneman have made popular, people are '*loss averse*', a loss of $\$x$ causes more pain than a gain of $\$x$ and they do not act rationally using mathematically correct expectations. The fear of loss is one of themes in Ionesco's bizarre play *The Lesson*, where a young woman comes for a tutoring lesson: she can add with proficiency but cannot subtract. The mathematician doesn't come off very rational either: he winds up killing her.

Mathematicians are attracted to Platonism, of believing that their discoveries are all insights into the eternal true world of mathematical facts. This example, the discovery of negative arithmetic and its incorporation into our numerical and algebraic toolkit, shows us that we must not be too literal. Yes, negative numbers were eventually accepted in the West as well as in China and India and all three cultures made the same math out of them. But there can be huge differences between cultures in the way mathematics unrolls. Euclid led the West down a certain path, dominated for many centuries by geometric figures and constructions. Other cultures were more practical and looked to solving concrete problems with approximate numbers. I think the discovery of calculus is another instance of this split: in India, studying the numerical table of sines led mathematicians to the idea of first and second differences and the fundamental theorem of calculus. But that is another story.

David Mumford
Professor Emeritus, Brown University
Chern Visiting Professor
UC Berkeley, Spring 2010
Email: `David_Mumford@brown.edu`

Kuṭṭaka, Bhāvanā and Cakravāla

Amartya Kumar Dutta

Ancient Indian mathematical treatises contain ingenious methods for finding integer solutions of indeterminate (or Diophantine) equations. The three greatest landmarks in this area are the *kuṭṭaka* method of Āryabhaṭa for solving the linear indeterminate equation $ay - bx = c$, the *bhāvanā* law of Brahmagupta, and the *cakravāla* algorithm described by Jayadeva and Bhāskara II for solving the quadratic indeterminate equation $Dx^2 + 1 = y^2$. We shall briefly recall the history of the above equations in ancient India and of their rediscovery in Europe, give an account of the ancient Indian texts dealing with algebra in general and the above indeterminate equations in particular, and mention a few works on history of mathematics which have highlighted ancient Indian algebra, especially indeterminate equations. We shall then discuss various mathematical aspects of the *kuṭṭaka*, *bhāvanā* and *cakravāla* from the viewpoint of modern algebra and number theory and the general cultural atmosphere in which the leading Indian mathematicians undertook such explorations. We shall also examine the coverage of these results in texts involving history of mathematics.

Introduction

In a plenary talk at the International Congress of Mathematicians (1978), André Weil[1] asserted ([We1, p 231–232]):

> An understanding in depth of the mathematics of any given period is hardly ever to be achieved without knowledge extending far beyond its ostensible subject-matter. More often than not, what makes it interesting is precisely the early occurrence of concepts and methods destined to emerge only later into the conscious mind of mathematicians; the historian's task is to disengage them and trace their influence or lack of influence on subsequent developments.

Indeed, one of the tendencies in recent research on history of mathematics has been to discover in ancient works the equivalents of modern concepts and ideas lying concealed in archaic language and style. For instance, I.G. Bashmakova's innovative studies[2] on the work of Diophantus, with this approach, have contributed

[1] André Weil (1906–98), one of the giants of 20th century mathematics, noted for his outstanding contributions in number theory and algebraic geometry, had a lifelong interest in the history of mathematics.

[2] A retrospective of the life and work of Isabella G. Bashmakova was published in Historia Mathematica **8** (1981), p 389–392, on the occasion of her 60th birthday; a more detailed and updated account,

to the revival of interest in Diophantus and enhanced the understanding of his work. However there is a dearth of literature which makes an analogous sophisticated analysis of ancient Indian works on indeterminate equations[3] (and other topics) in the light of modern mathematics.

Although the basic facts regarding ancient Indian works on the indeterminate equations $ay - bx = c$ and $Dx^2 + 1 = y^2$ are well-known among historians of Indian mathematics, not many mathematicians, even in India, are aware of these works. Consequently, there is a general lack of awareness about the heights attained in ancient Indian mathematics among scholars in other relevant disciplines (science, history, Indology, etc) as well as among general readers. Further, in several accounts on history of mathematics, one often finds misleading remarks about these works. Some of the accounts also reveal inaccurate notions on general Indian history.

In view of the above observations, the present paper has been prepared with the following aims:

(i) to present a mathematically non-technical history of the equations which will be accessible to readers who are not necessarily familiar with advanced mathematics;

(ii) to initiate a study on the significance of the *kuṭṭaka*, *bhāvanā* and the *cakravāla* in the light of modern algebra and number theory and to explore their pedagogic potential;

(iii) to recall relevant features of the cultural history of India during the "Classical Age";

(iv) to discuss some issues regarding the presentation of Indian works on indeterminate equations in texts on history of mathematics.

Roughly, sections 1–4 correspond to (i); sections 5–7 to (ii), and section 8 to (iii). Section 9 is specific to (iv); but the preceding sections, especially 3, 8 and parts of 5–7 are all relevant to (iv).

Section 1 gives a brief history of the methods for solving the equations $ay - bx = c$ and $Dx^2 + 1 = y^2$ in ancient India and in Europe. Section 2 gives an account of ancient Indian texts which deal with the equations while section

---

with bibliography of her writings, was published in Historia Mathematica **29** (2002), p 370–382, on her 80th birth anniversary.

[3]An algebraic (i.e., polynomial) equation in more than one variable (or more generally, a system of m algebraic equations in more than m variables) is called *indeterminate*. The term is suggestive of the fact that such a system may have infinitely many solutions. Diophantus of Alexandria (c. 250 CE) investigated rational solutions (solutions in rational numbers) of indeterminate equations with integer coefficients. In his honour, the indeterminate equations with integer coefficients whose solutions are to be found in integers (or sometimes in rational numbers) are called *Diophantine equations*, the adjective "Diophantine" pertains not so much to the nature of the equation as to the nature of the admissible solutions of the equation.

3 gives a partial account of the representation of Indian works on indeterminate analysis in texts on history of mathematics. Section 4 presents the history of development of certain topics of basic mathematics relevant to the *kuṭṭaka*. Sections 5–7 make a mathematical discussion on the *kuṭṭaka*, *bhāvanā* and *cakravāla* respectively. Section 8 highlights some features of the intellectual life of post-Vedic ancient India which are reflected in the legacy of Brahmagupta and other algebraists. Section 9 makes a critical discussion on certain views regarding the motivation of Brahmagupta and others in investigating the equation $Dx^2 + 1 = y^2$.

Readers interested in mathematical discussions may go directly to sections 5–7. Readers preferring non-technical discussions can first look at 1–3, 8, 9 and parts of 4. Parts of sections 5–7 (especially section 6) will also be accessible to general readers; they can go through these sections by skipping a few sentences involving mathematical technicalities.

1. A Brief History of The Problems

Indeterminate Analysis in Post-Vedic Ancient India

The mathematical treatises of the Classical Age in ancient India exude the general intellectual robustness of the era. Two outstanding works of the period are: *Āryabhaṭīya* of Āryabhaṭa (499 CE) and *Brāhma Sphuṭa Siddhānta* of Brahmagupta (628 CE). These texts had a decisive influence on the course taken by Indian mathematics (and astronomy) for the next 1000 years. In particular, they were responsible for the emergence of a flourishing algebra culture in Indian mathematics.

Apart from developing fundamental concepts and results of symbolic algebra[4], the algebraists of the Classical Age revelled in the challenging problems of finding integer solutions of indeterminate equations. The first explicit description of the general integral solution of the linear Diophantine equation $ay - bx = c$ occurs in the *Āryabhaṭīya* (499 CE) of Āryabhaṭa. The algorithm is described in just two cryptic stanzas (verses 32 and 33) at the end of the chapter *Gaṇita* (mathematics). The solution was subsequently discussed, with variations and refinements, by several Indian mathematicians including Bhāskara I (c. 600 CE)[5], Brahmagupta (628 CE), Mahāvīra (850 CE), Govindasvāmin (c. 860 CE), Pṛthūdakasvāmin (c. 860 CE),

[4]Algebra, as commonly understood, begins with the use of symbols (letters of the alphabet to denote unknowns) and equations ([DS2, Preface]). This phase of the subject is sometimes called symbolic algebra — the adjective "symbolic" is used to distinguish it from ancient geometric and arithmetical works where algebraic knowledge and ideas are only implicit.

[5]Different dates have been given for Bhāskara I like 522 CE (B. Datta), 574 CE (Sarasvati Amma) and 629 CE (K.S. Shukla). Brahmagupta's ninth-century commentator Pṛthūdakasvāmin places Brahmagupta (628 CE) to be later than Bhāskara I. We adopt 600 CE as a possible approximate date. See [Sar, p 9].

Āryabhaṭa II (950 CE), Śrīpati (1039 CE), Bhāskara II (1150 CE) and Nārāyaṇa (c. 1350 CE). The algorithm was termed *kuṭṭaka* (pulverizer), derived from the root *kuṭṭ* (to crush, to grind); the equation itself (or the problem of finding integer solutions of the equation) was also referred to as *kuṭṭaka*. The *kuṭṭaka* principle was considered so important that the subject algebra itself was initially called *kuṭṭaka-gaṇita*, or simply, *kuṭṭaka*. Bhāskara I gives 30 examples in his commentary *Āryabhaṭīya-bhāṣya* to illustrate the method for solving linear indeterminate equations (see [Shu4, p 309–334] and [Ke, 130–166] for the examples); examples are also given in his astronomy treatises *Mahā-Bhāskarīya* (see [Shu2, p 29–46]) and *Laghu-Bhāskarīya* (see [Shu3, p 99–114]). The *Mahāsiddhānta* of Āryabhaṭa II contains a separate chapter (18) called *kuṭṭaka* exclusively on the solution to the linear Diophantine equation. Bhāskara II discusses the *kuṭṭaka* in *two* of his treatises *Līlāvatī* ([PNS, chapter 33, p 167–175]) and *Bījagaṇita* ([Ab, chapter 5, p 17–21]; [P, chapter 5, p 49–69]). Devarāja wrote a separate treatise *Kuṭṭākāra-śiromaṇi* exclusively on the topic ([DS2, p 88]).

Having successfully dealt with the linear Diophantine equation, Indian algebraists took up the harder problem of investigating the quadratic indeterminate equation $Dx^2 + m = y^2$, called *varga-prakṛti* (square-natured) in ancient India. They laid special emphasis on solving the important case $Dx^2 + 1 = y^2$, where D is a positive integer which is not a perfect square.[6] The first major breakthrough in this problem was achieved by Brahmagupta in 628 CE through a brilliant innovation — a law of composition called *bhāvanā*. This composition principle of Brahmagupta is of paramount significance in modern algebra and number theory.

Brahmagupta applied his composition rule to generate an infinite number of integer solutions from a given integer solution of $Dx^2 + 1 = y^2$ and also showed how to arrive at a given integer solution in a wide variety of cases (that is, for various values of D like $D = 83$ or $D = 92$). While he had thus given a partial solution to the problem, another Indian algebraist subsequently discovered the complete integer solution by a cyclic method called *cakravāla* (*cakra*: wheel or disc) in ancient India. The discovery took place sometime during the 7th–11th century; for the *cakravāla* algorithm has been described by Jayadeva (who lived prior to 1073 CE)[7] and the famous astronomer-mathematician Bhāskara II (1150 CE). However, the true originator is not known (see next section). Bhāskara II illustrated the *cakravāla* with

[6]If D is negative, or if D is a square of a positive integer, then $Dx^2 + 1 = y^2$ has only finitely many integer solutions; in fact, it is easy to see that $(0, \pm 1)$ are the only solutions in all these cases except $D \neq -1$, and $(\pm 1, 0)$; $(0, \pm 1)$ are the only solutions when $D = -1$. It is when $\sqrt{D}$ is an irrational real number that the problem becomes mathematically interesting, and it turns out that the equation will have infinitely many solutions.

[7]Jayadeva's verses on the solution of the indeterminate equation $Dx^2 + 1 = y^2$ have been quoted in the text *Sundarī* of Udaydivākara composed in 1073 CE. Nothing is known so far about this mathematician Jayadeva. He is not to be confused with the Vaiṣṇava poet Jayadeva of the 12th century who composed *Gīta-Govinda*.

difficult numerical cases like $D = 61$ and $D = 67$. For the equation $61x^2 + 1 = y^2$, the smallest solution in positive integers is $x = 226153980$, $y = 1766319049$, indicating the unexpected intricacy of the problem.[8] Nārāyaṇa (c. 1350 CE) too discussed solutions of the equation $Dx^2 + 1 = y^2$, illustrated the method with the cases $D = 97$ and $D = 103$, and showed how to use the solutions to give rational approximations to $\sqrt{D}$ (see section 7). Besides the equations $ay - bx = c$ and $Dx^2 + m = y^2$, several other indeterminate and simultaneous indeterminate equations also occur in the treatises of some of the above mathematicians, especially Bhāskara II.

The equations $ay - bx = c$ and $Dx^2 + 1 = y^2$ are important equations in modern mathematics. But the Indian works on such indeterminate equations during the 5th–12th centuries were too advanced to be appreciated or noticed by Arab and Persian scholars and did not get transmitted to Europe during the medieval period. For instance, when Fyzi translated Bhāskara II's treatise *Līlāvatī*[9] into Persian, he omitted the portion on indeterminate equations.

We now narrate some of the landmarks in the history of the two indeterminate equations $ay - bx = c$ and $Dx^2 + 1 = y^2$ in post-Renaissance Europe.

Indeterminate Analysis in Modern Europe

Problems in indeterminate equations were taken up in Europe during the 17th century. These problems fascinated some of the best European mathematicians of the 17th–18th century who not only rediscovered solutions to the problems but also developed general theories and techniques in this connection.

The integer solution of the linear Diophantine equation was described in Europe for the first time by C.G. Bachet de Méziriac (1581–1638) more than eleven centuries after Āryabhaṭa. In the first edition (1612 CE) of his book *Problèmes plaisants et délectables qui se font par les nombres*[10], Bachet stated that if a and b are relatively prime integers, then one can find a least integral multiple of a which exceeds an integer multiple of b by a given integer c. The proof was given in the second edition of the book (1624). Dickson mentions that Bachet "employed notations for 18 quantities, making it difficult to hold in mind the relations between them and so obtain a true insight into his correct process" ([Di, p 44]). Bachet also mentioned

---

[8] Contrast it with the minimum solution of $60x^2 + 1 = y^2$: it is $x = 4$, $y = 31$.

[9] A popular text on arithmetic and mensuration with some algebra. It was translated by Fyzi from Sanskrit into Persian, at the behest of Emperor Akbar, in 1587 CE.

[10] *Pleasant and delectable problems to be solved by numbers*. Bachet had been interested in mathematical recreations and puzzles. This book is a collection of such puzzles. In 1621, Bachet published a bilingual edition (Greek original with Latin translation) of Diophantus' *Arithmetica* with extensive commentary. This work was largely responsible for the resurgence of number theory in Europe. Weil remarks ([We2, p 33]): "*. . . his is the merit of having provided his successors and notably Fermat with a reliable text of Diophantus along with a mathematically sound translation and commentary.*"

his solution in his edition (1621) of Diophantus' *Arithmetica* emphasising that the method is his own ([We2, p 7]). After Bachet's work, several mathematicians in Europe continued to discuss various approaches to the equation $ay - bx = c$. Some of the early names include J. Kersey (1673), M. Rolle (1690), T.F. de Lagny (1697), L. Euler (1735), N. Saunderson (1740), W. Emerson (1764), J.L. Lagrange (1767), J. Bernoulli (1772) among many others; see [Di, Ch II]. As late as in 1798 (nearly two centuries after Bachet), the great Legendre wrote in the preface of his *Theory of Numbers* (quoted in translation in [G, p 3]):

> Bachet,..., solved the indeterminate equation of the first degree by a general and very ingenious method.

Pierre de Fermat (1601–65), who is regarded as the father of Modern Number Theory, tried to create enthusiasm for the subject among his contemporary mathematicians like John Wallis (1616–1703) and Lord William Brouncker (1620–84) of England, B. Frenicle de Bessy (c. 1612–75) of France and F. van Schooten[11] (1615–60) of Netherlands. Fermat, who had carefully read Bachet's version of Diophantus, laid special emphasis on methods for finding *integer* solutions of equations as distinct from the easier, though important, question of finding rational solutions that occurs in Diophantus. One of the problems through which Fermat chose to highlight the beauty and intricacy of number theory was the problem of finding integer solutions to $Dx^2 + 1 = y^2$, a problem explored by Brahmagupta in 628 CE and solved by an Indian algebraist prior to the 11th century. Among the two specific examples that Fermat proposed to Frenicle in 1657 CE was the case $D = 61$ which occurs in Bhāskara II's *Bījagaṇita* (1150 CE).[12] To get some feel for the intensity of Fermat's involvement with number-theoretic methods, especially the equation $Dx^2 + 1 = y^2$, we may cite excerpts from his letter (1657) presenting his challenge to the English mathematicans (translation by Heath; quoted in [G, p 17]):

> There is hardly anyone who proposes purely arithmetical questions, hardly anyone who understands them. Is this due to the fact that up to now arithmetic[13] has been treated geometrically rather than arithmetically? This has indeed been the case both in ancient and modern works; even Diophantus is an instance. For, although he has freed himself from geometry a little more than others have in that he confines his analysis to the consideration of rational numbers, ...
>
> Now arithmetic has, so to speak, a special domain of its own, the theory of integral numbers. This was only lightly touched upon by Euclid in his Elements, and was not sufficiently

[11] A Dutch mathematician who is most famous for popularising the cartesian geometry of Descartes.

[12] The other case proposed by Fermat was $D = 109$. The smallest solution of $109x^2 + 1 = y^2$ is given by $x = 15140424455100$, $y = 158070671986249$.

[13] Here "arithmetic" refers to "higher arithmetic", i.e., number theory, the study of abstract properties of numbers. In Greece, usual computational arithmetic was called "logistic" while the term "arithmetic" meant number theory.

> studied by those who followed him . . . ; arithmeticians have therefore now to develop it or restore it.
>
> To arithmeticians therefore, by way of lighting up the road to be followed, I propose the following theorem to be proved or problem to be solved. If they succeed in discovering the proof or solution, they will admit that questions of this kind are not inferior to the more celebrated questions in geometry in respect of beauty, difficulty or method of proof.
>
> *Given any number whatever which is not a square, there are also given an infinite number of squares such that, if the square is multiplied into the given number and unity is added to the product, the result is a square.*

Regarding Fermat's challenge, André Weil remarked ([We2, p 81–82]):

> What would have been Fermat's astonishment if some missionary, just back from India, had told him that his problem had been successfully tackled there by native mathematicians almost six centuries earlier!

A general method for solving Fermat's problem was discovered by Brouncker in 1657.[14] During 1657–58, there had been an exchange of letters among mathematicians like Frenicle, Brouncker and Wallis who had taken interest in Fermat's problem. These letters were published by Wallis in *Commercium Epistolicum* (1658). In 1685, Wallis published his monumental work *A treatise of algebra both historical and practical* (in short, *Algebra*) comprising 100 chapters. Chapter 98 of this treatise was devoted to Fermat's problem and was based on the correspondence published in *Commercium Epistolicum*. Brouncker's method for solving the equation $Dx^2 + 1 = y^2$ was described in this chapter in a standard form. An enlarged second edition of the treatise *Algebra* was published as the second volume of Wallis' *Opera Mathematica* (1693).

Wallis had attempted to prove that $Dx^2 + 1 = y^2$ always has a positive integral solution, but used an incorrect result ([Di, p 354]). Fermat had asserted in his correspondences of 1659 that he had proved by his method of "descent" that the equation $Dx^2 + 1 = y^2$ has infinitely many integer solutions (when D is a positive integer which is not a perfect square). The proof has not been found in any of his writings.

Fermat's problem was again taken up in the next century by the two greatest figures of 18th century mathematics: L. Euler (1707–83) and J.L. Lagrange (1736–1813). Euler became interested in the problem from around 1730 and Lagrange took it up around 1768. They developed the theory of continued fractions, investigated the problem in the framework of this theory and established that if D is not a perfect square, then $\sqrt{D}$ has an infinite but periodic continued fraction expansion, and all

---

[14] In response to a challenge posed by Frenicle, Brouncker was able to use his method to find the (smallest) solution x=1819380158564160, y=32188120829134849 to the equation $313x^2 + 1 = y^2$ which he claimed had taken him only "an hour or two" ([OR]).

solutions (p, q) to the equation $Dx^2 + 1 = y^2$ are given by certain convergents $\frac{q_n}{p_n}$ of the expansion.[15] Moreover, there exists a positive integral solution (x_1, y_1) of $Dx^2 + 1 = y^2$ (called the *fundamental solution*) such that all positive integral solutions are given by (x_n, y_n) where x_n, y_n are defined by the relation $y_n + \sqrt{D}x_n = (y_1 + \sqrt{D}x_1)^n$. While the initial discoveries were made by Euler, it was Lagrange who first published formal proofs of all these results during 1768–69. He included them in his *Additions to Euler's Elements of algebra* (composed in 1771, published in 1773). Brouncker's method could be re-interpreted in the language of continued fractions; in that presentation, it becomes equivalent to Euler's method. The precise historical details are given in [We2, p 183; 229–233].

The Label "Pell's Equation"

Selberg mentioned in an interview[16] that André Weil once said that if something in mathematics gets attached to the name of a person, then the person in question usually has very little to do with it. The equation $Dx^2 + 1 = y^2$ was attributed to the English mathematician John Pell (c. 1611–85) by Euler although there is no evidence that John Pell had seriously investigated the equation. Euler's confusion could have been created through a cursory reading of Wallis' *Algebra*, a large part of which was devoted to the work of five English mathematicians: Oughtred, Harriot, Pell, Newton and Wallis himself. In any case, the name "Pell's equation" stuck. As Weil put it ([We2, p 174]):

> Pell's name occurs frequently in Wallis's *Algebra*, but never in connection with the equation $x^2 - Ny^2 = 1$ to which his name, because of Euler's mistaken attribution, has remained attached; since its traditional designation as "Pell's equation" is unambiguous and convenient, we will go on using it, even though it is historically wrong.

As suggested by R. Sridharan ([Sr2, p 17]), the equation should be called "Brahmagupta equation" as a tribute to the genius who first made a contribution to the problem, a thousand years before the time of Fermat and Pell.

Some Ancient Problems Involving $Dx^2 \pm 1 = y^2$

Some of the rational approximations to $\sqrt{D}$, which appear in (late Vedic) Indian and Greek texts, implicitly involve equations of the type $Dx^2 + 1 = y^2$. In the

---

[15] A well-motivated student-friendly account of the theory is presented in chapter 4 of ([G]). Students may also look at ([BrC, Chapters XXIV and XXXIII]) for an organised presentation of the theory of continued fractions with application to the equation $Dx^2 + 1 = y^2$.

[16] Bulletin AMS 45(4), October 2008, p 621.

Śulba-sūtras (c. 800 BCE), the most ancient extant mathematics treatises, $\frac{7}{5}$, $\frac{17}{12}$, $\frac{577}{408}$ have been used as approximations for $\sqrt{2}$ ([D2, p 202]; [Di, p. 341]; [Sar, p 18]). These three fractions may be interpreted as arising out of solutions of $2x^2 \pm 1 = y^2$; for $2 \times 5^2 - 1 = 7^2$, $2 \times 12^2 + 1 = 17^2$ and $2 \times 408^2 + 1 = 577^2$. In fact, they are respectively the third, fourth and eighth convergents of the simple continued fraction expansion of $\sqrt{2}$. We shall revisit these fractions in section 7.

Archimedes (287–212 BCE) gave the approximations $\frac{265}{153}$ and $\frac{1351}{780}$ for $\sqrt{3}$. In his commentary on the work of Archimedes, Eutocius (c. 480–540 CE) mentions the relations $265^2 - 3 \times 153^2 = -2$ and $1351^2 - 3 \times 780^2 = 1$ as a verification of the validity of the approximations ([We2, p 16]).

In 1773, G.E. Lessing discovered an ancient Greek epigram which states a problem (known as the "Cattle Problem") which is believed to have been communicated by Archimedes to Eratosthenes for the mathematicians of Alexandria. In this problem one is required to find eight integers (number of bulls and cows each of four colours) satisfying nine conditions linear and quadratic. After some algebraic manipulations, the cattle problem reduces to the problem of finding a positive integral solution to $Dx^2 + 1 = y^2$ where $D = 4729494$. There is no evidence that Archimedes had made this connection ([OR]) and it is thought extremely unlikely that either Archimedes or any of his contemporaries had actually solved the cattle problem. The problem was first solved by A. Amthor (1880 CE); the smallest value of one of the variables in the cattle problem is a number having 206545 digits (see [Di, p 342–345], [L] and [Wi, p 400–402] for further details).

Cakravāla Revisted

The Indian *cakravāla* needs fewer steps than the later methods of Brouncker and Euler; however a formal justification for the *cakravāla* method is less elegant. Some mathematicians therefore prefer to describe the Brouncker-Euler algorithm when writing on the *cakravāla* for students (see, for instance, [Vr, p 28–35]).

A formal proof that *cakravāla* method will always work (that is, terminate after a finite number of steps) was given by Krishnaswamy Ayyangar in 1929 ([Kr1]). Ayyangar related the method with the theory of a half-regular continued fraction[17] which he developed using only the kind of elementary mathematics which would have been known in India by the 5th century CE.

Another approach was given by C.O. Selenius (1960) who developed a generalised continued fraction expansion, which he called "ideal expansion", used it to

[17] See [Kr1], [Kr2], [Kr3] for Ayyangar's work and [MRW] for recent applications of Ayyangar's theory of "nearest square continued fraction".

give an algorithm for solving the equation $Dx^2 + 1 = y^2$ in which the number of steps is minimised, and showed that his "ideal expansion" method is equivalent to the Indian *cakravāla* ([Sel]).[18]

Legacy of "Pell's Equation"

The study of indeterminate equations, especially the study of the so-called "Pell's equation", played an important role in the evolution of classical algebra in ancient India and later in modern Europe.[19]

Pell's equation has had applications throughout history. Large integer solutions of $Dx^2+1 = y^2$ have been used in ancient times to yield good rational approximations to $\sqrt{D}$ (see section 7). The solutions of the equation yield units in the domain of integers of the quadratic field $\mathbb{Q}(\sqrt{D})$. The equation is also closely related to the study of binary quadratic forms.

The solution of Pell's equation is the main step in the solution of the general quadratic Diophantine equation in two variables. It also played a role in the solution (1970) of Hilbert's 10th problem on the non-existence of an algorithm for solving arbitrary Diophantine equations.

Pell's equation continues to fascinate mathematicians even today. Research papers and articles are being published on it in various contexts; two of the recent papers ([Ha], [MRW]) are listed in the bibliography. An extensive bibliography on the equation and on methods for solving it can be found in the 2002 paper ([Wi]) of Hugh Williams. The algorithmic efficiency of various methods are discussed in the article of H.W. Lenstra ([L]). An expository book ([Br]), exclusively on the topic, has been published recently (2003). The author E.J. Barbeau remarks in the Preface:

> Pell's equation seems to be an ideal topic to lead college students, as well as some talented and motivated high school students, to a better appreciation of the power of mathematical technique.[20]

Efficient generation of solutions of the equation is a very active area of research in algorithmic number theory and computer science. While there does not exist any polynomial-time algorithm for solving Pell's equation, S. Hallgren has exhibited a

---

[18] Selenius presented this work in a session on "Algebra and Theory of Numbers" at the International Congress of Mathematicians (1962). His abstract begins with a reference to Brahmagupta and Bhāskara ("Abstracts of Short Communications" of the ICM, Stockholm, p 51).

[19] I.G. Bashmakova has emphasised that solving indeterminate equations was a significant part in the development of algebra in Europe. See her article *Diophantine equations and the evolution of algebra* in the Proceedings of the International Congress of Mathematicians (ICM), Berkeley, California (1986).

[20] This is also the view of the present author who feels that the Indian works are particularly suited for this purpose (see [Du1, p 86–89]).

polynomial-time quantum algorithm for the equation ([Ha]). As Lenstra remarks ([L, p 192]):

> The last word on algorithms for solving Pell's equation has not been spoken yet.

2. Texts on Algebra in the Classical Age

As mentioned earlier, the subject of algebra was developed during the Classical Age, consciously, as a distinct and important branch of mathematics. The text *Āryabhaṭīya* (499 CE) already contains algebraic ideas, like a term *gulikā* ("shot"; possibly "coloured shot") for an unknown quantity and algebraic results like solutions of linear and quadratic equations, formulae for arithmetic progression and other sums of finite series and, of course, the integer solutions of linear Diophantine equations. Bhāskara I (c. 600 CE) discusses the *kuṭṭaka* (the linear Diophantine equation) in detail in the first chapter (verses 41–52) of his astronomy treatise *Mahā-Bhāskarīya* ([Shu2, p 29–46]), apart from the discussions, with examples, in his commentary *Āryabhaṭīya-bhāsya* ([Ke, 128–166]).

It is however the *Brāhma Sphuṭa Siddhānta* (628 CE) of Brahmagupta which laid a firm foundation for the science of algebra in India. This is the first known ancient Indian text containing a separate chapter on algebra. Chapter 18 of this voluminous treatise (comprising over 1000 verses in 24 chapters) is titled *Kuṭṭakādhyāyaḥ* (= *Kuṭṭaka* (algebra) + *adhyāyaḥ* (chapter)). Towards the beginning and the end of this chapter, Brahmagupta emphasises the importance (rather indispensability) of the subject and its charm. He systematically expounds the basics: describes the use of symbols for unknowns, defines zero as an integer in algebra, introduces negative numbers and prescribes the rules of arithmetic operations in the enlarged number system (including negative numbers and zero), discusses the arithmetic of surds, explains the principles of formation and manoeuvring of equations (plan of writing, arrangement, clearances, etc), gives methods for solving quadratic equations and simultaneous linear equations in several unknowns (anticipating the Gaussian method of elimination), etc. A substantial portion of the chapter on algebra discusses indeterminate equations of the first and second degree.

The impact of Brahmagupta on algebra can be seen from the fact that separate monographs on the subject (usually titled *Bījagaṇita*[21]) were brought out subsequently by prominent mathematicians like Śrīdhara (750 CE), Padmanābha (date unknown but before Bhāskara II), Śrīpati (1039 CE), Bhāskara II (1150 CE),

[21] The subject of algebra gradually acquired the current Sanskrit term *bījagaṇita*. *Bīja* means "seed", "root", "element", as well as "analysis"; thus *bījagaṇita* literally means the "mathematics with elements (letters)" as well as "analytical mathematics".

Nārāyaṇa (1350 CE) and later mathematicians. Unfortunately, the specialised algebra texts of Śrīdhara, Padmanābha and Śrīpati — the predecessors of the great Bhāskara II — have not been found.[22]

Apart from algebra texts, the general texts on mathematics — like the *Gaṇita-sāra-saṅgraha* (750 CE) of Mahāvīra — also contain several results on topics in algebra like quadratic equations, summations of finite series (like arithmetic progression (AP) and geometric progression (GP)), permutations and combinations, and sometimes even the *kuṭṭaka*. Some of the astronomer-mathematicians wrote separate chapters on algebra in their general texts. Chapter 14 of the astronomy treatise *Siddhāntaśekhara* of Śrīpati, entitled *avyakta-gaṇita*, is on algebra.

Discussions on the relatively advanced or specialised topics in algebra, like solutions of quadratic indeterminate equations, were usually confined to the *Bījagaṇita* texts. The *Bījagaṇita* of Bhāskara II contains — apart from basic results of algebra — a lucid exposition of the *kuṭṭaka*, *bhāvanā* and the *cakravāla*, and mentions several other indeterminate equations. Bhāskara II attributes the *cakravāla* to ancient authors without specifying any name. At the end of his treatise, Bhāskara II makes a general acknowledgement of the prolific algebra works of Brahmagupta[23], Śrīdhara and Padmanābha as his chief sources. There is therefore a strong possibility that the lost treatises of Śrīdhara and/or Padmanābha too contain an account of the *cakravāla* algorithm. The history of *cakravāla* became more mysterious with the discovery (in 1954) of a work dated 1073 CE which quotes verses from a preceding algebraist Jayadeva describing the *cakravāla* algorithm (along with the *bhāvanā*). Curiously, Jayadeva is not mentioned by Bhāskara II.

It is remarkable how early the leading Indian mathematicians had realised the significance of algebra and how strongly they asserted and established the importance of their newly-founded discipline. Brahmagupta declares at the outset of his *Kuṭṭakādhyāyaḥ* (chapter on algebra) that algebra is indispensable for solving most problems, gives a list of topics in algebra, and declares that one who has mastered those topics will be venerated as an *ācārya* among the learned. And, towards the end of the chapter, he celebrates the delight and glory of the subject through the words ([Sha, Ch 18; verses 99–100]; [Du1, p 104]):

> These questions are stated simply for delight. One may devise a thousand others, or may resolve the problems proposed by others, by the rules given here.
>
> As the sun eclipses the stars by his brilliance, so will an expert eclipse the glory of other astronomers in assemblies of people by the recital of algebraic problems, and still more by their solution.

---

[22]Bhāskara II refers to the extensive algebra works of Śrīdhara and Padmanābha ([Ab, p 52]; [P, p 196]); Munīśvara (c. 1608 CE) mentions the *Bījagaṇita* of Śrīpati ([Sih, p 26]).

[23]The verse of Bhāskara II ([Ab, p 52]; [P, p 196]) actually mentions Brahmā. It is generally presumed by historians that Bhāskara II was referring to Brahmagupta.

In order to emphasise the importance, power, and profundity of algebra, the poet-mathematician Bhāskarācārya (Bhāskara II) begins his treatise *Bījagaṇita* with an Invocation involving an interesting "pun" on the words *Sāṁkhyāḥ* (the Sāṁkhya philosophers as well as the experts in *saṁkhyā*, the science of numbers), *Satpuruṣa* (the Self-Existent Being as well as the wise mathematician), *bīja* (root/cause as well as algebra) and *vyakta* (the manifested universe as well as the revelation of an unknown quantity). Thus, through the opening verse, Bhāskarācārya venerates the Unmanifested — the Self-Existent Being of the Sāṁkhya philosophy — who is the originator of intelligence and the primal Cause of the known or manifested universe; and, through the very same words, Bhāskarācārya pays tribute to the wise mathematician who, using algebra, solves a problem (i.e., reveals or manifests an unknown quantity). The importance of algebra is reiterated at the end of *Bījagaṇita*. Bhāskarācārya remarks that algebra is the essence of all mathematics, is full of virtues and free from defects; and that the cultivation of algebra will sharpen the intellect of children. He concludes with the exhortation "*paṭha paṭha*" (Learn it, learn it) for the development of intelligence ([Ab, p 7; 53]; [P, p 4; 198]).

Another algebraist Nārāyaṇa (1350) exalts the subject with the following tribute ([DS2, p 5]):

> As out of Him is derived the entire universe, visible and endless, so out of algebra follows the whole of arithmetic with its endless varieties (of rules). Therefore, I always make obeisance to Śiva and also to (*avyakta-*) *gaṇita* (algebra).

3. Ancient Indian Algebra in History of Mathematics

In this section, we give a brief account of the discovery of some of the above works by modern historians and of expositions on them in some of the important early texts on history of mathematics.

English translations of ancient Indian mathematics texts, especially of Bhāskara II, began appearing from early 19th century. Bhāskara's *Bījagaṇita* (algebra) was translated from Persian[24] into English by E. Strachey under the title *Bija Ganita or the Algebra of the Hindus* (London, 1813). The *Līlāvatī* of Bhāskara was translated from the original Sanskrit by John Taylor (Bombay, 1816).

The work during this period which, perhaps, had the greatest impact in generating an awareness about ancient Indian mathematics was the translation by H.T. Colebrooke titled *Algebra, with Arithmetic and Mensuration, from the Sanscrit of Brahmegupta and Bhascara* (London; 1817). Colebrooke translated the texts *Līlāvatī* and *Bījagaṇita* of Bhāskara II and the chapters *Gaṇitādhyāyaḥ* (Ch 12; arithmetic and mensuration) and *Kuṭṭakādhyāyaḥ* (Ch 18; algebra) of *Brāhma Sphuṭa Siddhānta*

---

[24]Based on the Persian translation of 1634 CE by Ata Alla Rushdi.

of Brahmagupta, along with selections from the reputed commentary on *Brāhma Sphuṭa Siddhānta* titled *Vāsanābhāṣya* (c. 860 CE) by Caturveda Pṛthūdakasvāmin.

Colebrooke also wrote a long preface where he discussed some of the distinguishing features of the texts. This preface was reproduced in Colebrooke's *Miscellaneous Essays* (1837) and its contents thereby got the attention of a wide community of serious scholars on India's cultural history. One of the features of Indian algebra that Colebrooke highlighted was ([Co2, p 438]):

> General methods for the solutions of indeterminate problems of first and second degrees, in which they went far, indeed, beyond Diophantus, and anticipated discoveries of modern algebraists.

For some time during the 17th–19th centuries, *Āryabhaṭīya* had been unavailable to scholars of Indology, though its existence and contents had been known through other available texts. In 1817, Colebrooke laments ([Co2, p. 422]):

> A long and diligent research in various parts of India, has, however, failed of recovering any part of the *Padmanābha vīja* (or Algebra of Padmanābha[25]), and of the algebraic and other works of Āryabhaṭṭa.

Although a few manuscripts of *Āryabhaṭīya* were available in parts of Kerala, Indologists did not come across them for several decades till Bhau Daji[26] discovered a manuscript in 1864. The text, with the commentary by Parameśvara, was then published by H. Kern in 1874 at Leiden (Netherlands). Subsequently, there have been several English translations of *Āryabhaṭīya* with commentaries by historians of science like P.C. Sengupta from Calcutta (1927) and W.E. Clark from Chicago (1930). In the 1500th birth anniversary year of Āryabhaṭa (b. 476 CE), the Indian National Science Academy (INSA) published the *Āryabhaṭīya* with English translation and notes by K.S. Shukla and K.V. Sarma ([ShuS]), along with the commentaries of Bhāskara I and Someśvara (edited by K.S. Shukla) and the commentary of Sūryadeva (edited by K.V. Sarma). Recently, A. Keller has prepared an English translation of the portion of Bhāskara I's *Āryabhaṭīya-bhāṣya* which discusses the mathematics chapter *Gaṇita* of *Āryabhaṭīya* ([Ke]).

The text *Gaṇita-sāra-saṅgraha* (750 CE) of Mahāvīra was discovered in early 20th century and published in 1912 with an English translation by M. Rangacharya.

---

[25] The works of Padmanābha remain elusive.

[26] A very successful and popular medical practitioner, Dr. Bhau Daji (1821–1874) made immense contributions to Indological research: travelling throughout the country (visiting forts, temples and caves) in search of ancient inscriptions; collecting a large number of manuscripts and a good number of copper-plates and movable epigraphs; and publishing several papers on the history of Indian numerals, chronology, epigraphy and literature.

The works of Bhāskara I were discovered by B. Datta[27] in 1930 and published by K.S. Shukla during 1960–76 ([Shu2], [Shu3], [Shu4]).

In 1954, K.S. Shukla discovered the text *Sundarī* (1073 CE) of Udaydivākara, a commentary on the *Laghu-Bhāskarīya* of Bhāskara I, and found that it contains quotations from a hitherto unknown mathematician called Jayadeva on the *bhāvanā* and the *cakravāla*. He published the verses in [Shu1].

The book by Datta-Singh ([DS2]), published in 1938 and reprinted in 1962, remains the richest reference book on ancient Indian algebra. Much information on indeterminate analysis is recorded in the book. However there is not much discussion on the significance of the works from the viewpoint of modern algebra. The ideas of abstract modern algebra had not become commonplace among the general mathematical community in India in the 1920s and 30s when Datta did his pioneering work on history of mathematics.

A few books published from the West in early 20th century had incorporated facts on ancient Indian mathematics with reasonable accuracy. L.E. Dickson's *History of the Theory of Numbers* Vol II (1920) contains considerable mention of Indian contributions to the study of indeterminate equations. In fact, in the Preface, he makes a special mention of Brahmagupta's work ([Di, p. xi]):

> It is a remarkable fact that the Hindu Brahmegupta[28] in the seventh century gave a tentative method of solving $ax^2 + c = y^2$ in integers, which is a far more difficult problem than its solution in rational numbers.

However, Dickson was not aware of the discovery of *Āryabhaṭīya* in 1864 and, like Colebrooke, refers to it as a lost treatise ([Di, p 41]).

The book *History of Mathematics* by F. Cajori (1919) contains a short chapter "The Hindus" where there is a brief but admiring reference to Indian works on indeterminate equations ([Ca, p 94–98]):

> ... Indeterminate analysis was a subject to which the Hindu mind showed a happy adaptation. We have seen that this very subject was a favorite with Diophantus, and that his ingenuity was almost inexhaustible in devising solutions for particular cases. But the glory of having invented *general* methods in this most subtle branch of mathematics belongs to the Indians. The Hindu indeterminate analysis differs from the Greek not only in method, but also in aim. The object of the former was to find all possible integral solutions. Greek analysis, on the other hand, demanded not necessarily integral, but simply rational answers.

---

[27]Bibhutibhusan Datta (1888–1958), later Swami Vidyaranya, is a pioneer among historians of Indian mathematics. His numerous papers contain a wealth of information and insights for anyone seriously interested in the history of ancient Indian mathematics; unfortunately most of them are not easily accessible. A D.Sc. in Applied Mathematics, Datta possessed the rare combination of professional mathematical experience with deep knowledge of Sanskrit literature. Apart from his books on history of Indian mathematics ([D2], [DS1], [DS2]), he has also written a few books on Indian philosophy.

[28]The effect of Colebrooke on Dickson can be seen in the spelling of Brahmagupta.

> Diophantus was content with a single solution; the Hindus endeavored to find all solutions possible. . . .
>
> Remarkable is the Hindu solution of the quadratic equation $cy^2 = ax^2 + b$. With great keenness of intellect they recognized in the special case $y^2 = ax^2 + 1$ a fundamental problem in indeterminate quadratics. They solved it by the *cyclic method*. . . .
>
> Doubtless this "cyclic method" constitutes the greatest invention in the theory of numbers before the time of Lagrange.[29] . . .
>
> . . . Unfortunately, some of the most brilliant results in indeterminate analysis, found in Hindu works, reached Europe too late to exert the influence they would have exerted, had they come two or three centuries earlier.

While several pioneering scholars like Colebrooke, Hankel and Thibaut attempted objective presentations of different portions of ancient Indian mathematics, there were other influential writers, steeped in Eurocentric views, who sought to undermine it. The authors who were prejudiced in their portrayal of ancient Indian mathematics did not even bother to be accurate or careful in conforming to known facts. For instance, any work on the Pellian equation having some sophistication was believed to be of Greek origin; and many authors who favoured the hypothesis of Greek origin did not have clear ideas regarding the distinction between Brahmagupta's brilliant but partial work and the subsequent complete *cakravāla* algorithm (which is a strong indicator that the research was indigenous). Two errors commonly occurred: there were some who had the impression that the final *cakravāla* method also occurs in Brahmagupta's work; and there have been writers who thought that Bhāskara II did not go beyond Brahmagupta's partial work. We refrain from giving details in this article. While there has been a recognition of the bias of certain writers, their warped views and erroneous remarks have had an influence on several subsequent authors and result in a distorted representation of ancient Indian mathematics even by more objective writers. (See also section 9.)

Eurocentrism declined with passage of time. But in the subsequent period there has not been as much of dedicated scholarship on ancient Indian mathematics as compared to, say, Greek or Babylonian mathematics. Moreover, well-edited publications with English translations and commentaries of even the major ancient Indian mathematical treatises are not always available easily. Let us consider the current situation regarding some of the ancient texts relevant to the present article. The INSA publication on *Āryabhaṭīya* ([ShuS]) has long been out-of-print. As far as the present author is aware, the original verses of *Brāhma Sphuṭa Siddhānta*

---

[29] There seems to be an allusion to the remark on *cakravāla* by H. Hankel: "*It is beyond all praise; it is certainly the finest thing achieved in the theory of numbers before Lagrange.*" (Zur Geschichte der Mathematik in Alterthum und Mittelalter, Leipzig, 1874, p 202.)

were last published in 1966 ([Sha]), with Hindi translation. Not all chapters (not even all the mathematics chapters) of Brahmagupta's magnum opus have English translations. Colebrooke's partial translations ([Co1]), though valuable, are not always lucid. Jayadeva's verses occur only in the 1954 paper ([Shu1]). The picture is somewhat better with respect to the mathematics texts of Bhāskara II. A well-known publisher has brought out an English edition, with notes, of *Līlāvatī* ([PNS]), based on the Marathi work of N.H. Phadke. Two useful recent editions of *Bījagaṇita* are mentioned in the bibliography ([Ab]; [P]); however, neither is free from errors. Publications on ancient commentaries of the original treatises are even more scarce. As M.D. Srinivas emphasises in [Srn], commentaries are no less important than the original treatises for a proper comprehension of the history of ancient Indian mathematics.

Results like *bhāvanā* and *cakravāla* have evoked deep appreciation from mathematicians of high eminence, as can be glimpsed from the statements of Weil (sections 1,5,7) and Atiyah (section 9) or the tribute of Ojanguren (section 6). However, such observations occur at diffused places — as incidental remarks in a sophisticated book, in an isolated lecture, or in a highly specialised research monograph — far removed from the attention of the general scholar. Going by some of the books on history of mathematics in recent decades, it would appear that not all historians are sufficiently aware of how highly the ancient Indian works are regarded by some of the great mathematicians of modern times.[30]

4. Long Division, GCD and Euclidean Algorithm

Before commencing our discussion on the *kuṭṭaka*, we shall make a brief mention of the Euclidean algorithm which is a key step in the solution of the linear indeterminate equation. We first discuss the antiquity of the method of "long division" in ancient India.

Computational Arithmetic in Ancient India: Long Division

Familiarity with the four fundamental operations of arithmetic is evident in Vedic literature like the *Śatapatha Brāhmaṇa*, the *Taittirīya Saṁhitā* and even the

[30]Many accounts, even by highly respected historians of mathematics, omit Indian contribution altogether. To give two examples:- *The History of Mathematics: A Reader* (1987) by J. Fauvel and J. Gray has no representation of mathematics in India; the monograph *The Beginnings and Evolution of Algebra* (2000) by I.G. Bashmakova and G.S. Smirnova has no mention of the algebra of post-Vedic ancient India.

Ṛg-Veda. A passage from the ancient *Śatapatha Brāhmaṇa* (quoted in [Sr2, p 5–6]) gives all divisors of 720. The *Śulba-sūtra*s show clear use of arithmetic operations including division; they even give examples of division by fractions. We however do not know about the actual techniques used in the Vedic era.

Now, although there was a long tradition of computational mathematics in ancient India, a major difficulty in the reconstruction of this history is that the post-Vedic mathematical works preceding *Āryabhaṭīya* (499 CE) are lost.[31] Āryabhaṭa mentions at the beginning (verse 1) of the mathematics chapter *Gaṇita* of *Āryabhaṭīya* that he is recording ancient knowledge "honoured at Kusumapura". His commentator Bhāskara I makes an incidental reference to names like Maskarī, Pūraṇa, Mudgala, Pūtana, referring to them as *Ācārya*s (Masters).[32] However nothing is known to us about such revered mathematicians.

The gaps in our knowledge regarding the development of computational techniques in India may be illustrated by the history of "division". A variant of the modern method of "long division" is described in the arithmetic treatises of Śrīdhara (750 CE), Mahāvīra (850 CE) and later writers (see [DS1, p 150–154] for details). But how old is their method of "long division"? Āryabhaṭa's ingenious algorithms for computing square roots and cube roots (which are again slight variants of the present methods) involve the technique of long division. However, neither Āryabhaṭa (499 CE) nor Brahmagupta (628 CE) bother to explain any method of division separately. They take for granted the knowledge of the long division method when they describe their algorithms for extraction of square roots and cube roots. One can only conclude that the device had become firmly established well before the 5th century and that the great Masters Āryabhaṭa and Brahmagupta considered the method to be too elementary for their respective treatises.

We mention here that the operation of division was regarded as a difficult operation by European scholars from ancient times till as late as the 15th century. Luca Pacioli remarked in 1494 CE: "*if a man can divide well, everything else is easy, for all the rest is involved therein.*"([Sm, p 132]). Not much is known about ancient European methods of division. The Indian version of long division was transmitted to the Arabs and occurs in Arabic works from the 9th century onwards. From the Arabs, the method went to Europe where it came to be known as the *galley* (*galea*, *batello*) method (also called "scratch method") which became very popular in Europe during the 15th–18th centuries. The method was very convenient in the

---

[31] Perhaps the original text of the Bakṣālī manuscript could be anterior to *Āryabhaṭīya*. Its date is controversial; some historians like B. Datta place it at around 200 CE.

[32] Evidently they had composed voluminous mathematics texts; for Bhāskara I explains why Āryabhaṭa records only a "bit of mathematics" ([Sar, p 68–69]; [Shu4, p liv–lv]): "*This has not been composed as a treatise by ācāryas Maskarī, Pūraṇa, and others, with even one lakh verses for each [topic]. How can the Ācārya [Āryabhaṭa] manage [to state all of] it within such a short treatise?*"

earlier system where erasing of figures was easy and desirable.[33] The paper-and-ink system was less conducive for it and eventually led to the modern arrangement — a modification of the galley method.

Greatest Common Divisor (GCD)

As the name suggests, the GCD of two positive integers is the largest integer which divides both the numbers (without leaving any remainder). As we shall see in next section, the first major step in the *kuṭṭaka* is the computation of the GCD of two numbers. In the *Āryabhaṭīya*, the algorithm for computing GCD is alluded to in a terse way. In fact, all expositors on *kuṭṭaka* — even those who are generally more elaborate — refer to the algorithm for computing GCD by repeated division without much elaboration.

One can only conclude that, before the time of Āryabhaṭa, ancient Indians had become familiar with the technique of computing GCD, presumably by long division. As is the case with most concepts and results occurring in *Āryabhaṭīya*, we lack the precise knowledge regarding the genesis of the concept of GCD in ancient India.

Amidst all literature that is available to us, Euclid's *Elements* is the oldest text which describes an algorithm for computing the GCD. However, there is no concrete evidence that this algorithm was transmitted to Indian mathematicians. As one does not see any influence of other results from the *Elements* in the mathematics of Āryabhaṭa and his successors (see the discussion in the second paragraph of (iv) of section 9), it is more likely that a division method for computing GCD evolved independently in ancient India.

Reciprocal Subtraction

Let a and b be the lengths of two line segments (that is, two "real numbers" in modern terminology). If $a > b$, then one subtracts b from a, then b from $a - b$ if $a - b > b$, then b from $a - 2b$ if $a - 2b$ still exceeds b, and so on, till one reaches a stage, say after q_1 subtractions, when the remainder $r_1 := a - q_1 b$ does not exceed b. Now, if $b > r_1$, then, as before, one keeps on subtracting r_1 from b as often as possible, say q_2 times, till one reaches the remainder $r_2 = b - q_2 r_1$ which is $\leq r_1$. The subtractive process is then continued with the lengths r_1 and r_2 to get q_3, r_3, then with r_2 and r_3, and so on. This process was called *Antanairesis* (reciprocal subtraction) in ancient Greek mathematics. The sequence (finite or infinite) of quantities $\{q_1, q_2, \dots\}$ was

[33]In ancient India, calculations (*pāṭīgaṇita*) were usually performed on sand spread on a board called *pāṭī*. In this format, the figures being big, not many lines of figures could be written on a board; but it was easy to erase figures. Comuptational techniques, therefore, were planned in such a way that fewer figures would be used at a time, and figures which are no longer required could be erased. See [DS1, p 129].

used to define the concept of *Logos* (ratio) in the theory of proportions developed by Greek mathematicians preceding Eudoxus (400–349 BCE). For further details, see [F] and [Vn, p 90–91].

Euclidean Algorithm

Books VII, VIII and IX of Euclid's *Elements* ([He]) describe the basic multiplicative properties of positive integers. As Weil remarks ([We2, p 4]):

> It is generally agreed upon that much, if not all, of the content of those books is of earlier origin, but little can be said about the history behind them.

Irrespective of the actual genesis of specific results, the *Elements* (c. 300 BCE) is the oldest extant book which makes a systematic exposition of the fundamentals of "Number Theory".

The Euclidean algorithm for finding GCD is described at the very beginning of Book VII of the *Elements* (Propositions 1–2); and presented again, in a geometrical form, in Book X (Propositions 2–3).[34] We briefly recall the method; for clarity, we take advantage of modern language, especially the subscript notation. For a more detailed mathematical exposition of the Euclidean algorithm and associated concepts like GCD at an elementary level, the reader may refer to [JJ, Ch 1].

Note that, for integers, the principle of *Antanairesis* or mutual "subtraction" may be abbreviated to the following principle of "division": *Given two positive integers* a, b *with* $a \geq b$, *there exist unique non-negative integers* q, r *such that* $a = bq + r$ *with* $0 \leq r < b$. This result is often called the "Division Algorithm", although it is a theoretical statement rather than a practical algorithm — it does not describe any method to find q and r when a and b are given. Although Euclid too uses the language of repeated mutual subtraction in Propositions 1–2 of Book VII ([He, Vol 2, p 296–299]), the above principle (of taking the remainder of one number *divided* by another) is implicit in his proof. It is not known how Greeks performed division during Euclid's time,[35] and how they implemented the "Division Algorithm" for large numbers (by any method more efficient than repeated subtraction). As mentioned earlier, during Renaissance in Europe, the modern method of "Long Division" evolved out of the corresponding ancient Indian method.

---

[34] In Book VII, the algorithm is formulated for integers; in Book X, it is formulated for lengths of line segments. In the latter case, one defines the GCD, or rather its synonym the Greatest Common Measure (GCM), of two lengths a and b to be the greatest length g such that a and b are both integer multiples of g. The GCM will exist if and only if a and b are commensurable, that is, in our language, if and only if $\frac{a}{b}$ is a rational number.

[35] Smith writes ([Sm, p 133]): "*We are quite ignorant as to the way in which the Greeks and Romans performed the operation of division before the Christian Era.*"

By the so-called Division Algorithm (or by *Antanairesis*), given two positive integers a, b, one has a sequence of relations $a = a_1 b + r_1$, $0 < r_1 < b$; $b = a_2 r_1 + r_2$, $0 < r_2 < r_1$; $r_1 = a_3 r_2 + r_3$, $0 < r_3 < r_2$; and so on, till one arrives at a stage where $r_{n+1} = 0$, that is, $r_{n-1} = a_{n+1} r_n$. The Euclidean algorithm asserts that r_n is the GCD of a and b. It is based on the principle that the GCD of two numbers does not change if the smaller number is subtracted from the larger number. For instance, the GCD of 24 and 18 is same as the GCD of 6 (= 24 − 18) and 18. Consequently, $\text{GCD}(a, b) = \text{GCD}(a - b, b) = \cdots = \text{GCD}(a - a_1 b, b) = \text{GCD}(b, r_1) = \text{GCD}(r_1, r_2) = \cdots \text{GCD}(r_{n-1}, r_n) = r_n$.

Although *Elements* is the earliest available text which gives a clear description of the algorithm, scholars of Greek mathematics believe that it was known much earlier, at least in its subtractive form. The Euclidean algorithm was known in China and in India. As explained earlier, in the absence of enough historical documents, it is not known whether it was discovered independently and when the algorithm was first introduced or discovered in these countries.

The Euclidean algorithm is one of the oldest algorithms that is still in common use. A crucial feature of the algorithm is that it finds the GCD efficiently without requiring the determination of the prime factors of the numbers.[36] The algorithm is *logarithmic* in the size of integers. G. Lamé (1845) proved that the number of division steps required by the Euclidean algorithm to compute the GCD of two given integers never exceeds five times the number of digits (in decimal notation) of the smaller integer ([Kn, p 343]).

5. Kuṭṭaka

Brevity in Indian Tradition

The original mathematical treatises in ancient India (especially the earlier texts) were generally very brief. The verses were meant to indicate broad hints. There was emphasis on the use of a learner's own intellect for working out the complete details.[37] Detailed expositions on the works of the Masters were orally transmitted through the *guru-śiṣya* (mentor-disciple) link. Sometimes the commentaries give details. The approach of the ancient stalwarts can be seen from the following

[36]Factorisation of large integers is regarded as a difficult problem; in fact, many systems of modern cryptography are based on the non-availability of means to determine factorisation of sufficiently large integers.

[37]As an extreme example, consider the brevity of Āryabhaṭa's instruction for constructing a sphere which would rotate uniformly at the rate of Earth's rotation (Gola 22): "*[Make] a wooden gola (globe) which is perfectly spherical and uniformly dense all around [but is] light in weight. Using mercury, oil and water, and applying one's own intellect, [make the globe] rotate [at the required rate] to keep pace with time.*" See [Du2, Resonance **11**(3), p 67].

remarks that Bhāskara II makes in the epilogue of his text *Bījagaṇita* ([Ab, p 53]; [P, p 196–197]):

> The expanse of science is vast as the ocean, which a person of ordinary intellect finds too formidable to cross. But what is the necessity of details for the intelligent student? A mild instruction suffices; for it helps the intelligent student to develop the knowledge on his own. Just as a drop of oil put in water, or a grain of secret confided to a villain, or a little act of charity to the deserving, spreads automatically, likewise a quantum of knowledge, instilled into an intelligent mind, grows and expands extensively by its own force.

This attitude is not much different from that of most modern mathematicians who too feel that a budding researcher eventually gains more insight from a terse text than from a clearly spelt-out text, from an obscure important paper than from a lucid one. (For further discussion, see [DS1, p 126]; [Du1, Remark 4; p 110–111].)

The original description of *kuṭṭaka* by Āryabhaṭa (see Appendix) is excruciatingly brief. Historians of mathematics have interpreted his two verses based on ancient commentaries; for instance, B. Datta has been guided by the explanation of Bhāskara I, the earliest commentator on *Āryabhaṭīya* whose work is available to us. The descriptions by Bhāskara I, Brahmagupta and several other writers are quite similar, in essence. Āryabhaṭa II observed some simplifying devices which were emulated by later writers. In *Bījagaṇita*, Bhāskara II makes a neat presentation of the main algorithm and the useful auxiliary principles and states a few suggestive examples where one could implement the prescribed rules ([Ab, Ch 5]; [P, Ch 5]).

In this section, we shall make a summary of the various presentations of the *kuṭṭaka* by authors ranging from Bhāskara I to Bhāskara II. Details take up 54 pages in the book of Datta-Singh ([DS2, p 87–140]). We shall try to highlight the key ideas which would be particularly enriching for students of mathematics.

Formulation of the *Kuṭṭaka*

As the linear indeterminate equation had applications in astronomy, most Indian mathematicians considered the problem of determining all *positive* integral solutions of equations of the type $ay - bx = \pm c$, where a, b, c are positive integers.

As an illustration of how a concrete problem like $100y - 63x = \pm 90$ would be usually phrased in verses, we quote from the *Bījagaṇita* (also *Līlāvatī*) of Bhāskara II ([Ab, p 20]; [P, p 58]; [PNS, p 167]):

> If thou be an expert in *kuṭṭaka* analysis, please tell me that number(s) which, when multiplied by 100 and increased or decreased by 90, would become divisible by 63 without a remainder.

Thus the equation $ay - bx = \pm c$ was visualised in the form $y = \frac{bx \pm c}{a}$. The quantities a, b, c, x, y were called *hāra* or *bhājaka* (divisor), *bhājya* (dividend),

kṣepa (interpolator), *guṇaka* (multiplier) and *phala* or *labdhi* (quotient) respectively. Bhāskara I arranges the equation in the form $y = \frac{bx+c}{a}$ or $x = \frac{ay+c}{b}$ so as to ensure that the interpolator c comes with a positive sign. Almost all ancient Indian writers observe, in some form, that the equation $ay - bx = c$ will have integral solutions only if c is divisible by the GCD of a and b called *apavartana* or *apavarta*.

Reductions

The treatment of the *kuṭṭaka* reveals a tendency by ancient Indians to employ "reductions"[38]; an aspect which has pedagogic value for training budding mathematicians. We shall give a few examples.

First, Bhāskara I, Brahmagupta, Āryabhaṭa II, Śrīpati and Bhāskara II, among others, explicitly state that all the coefficients should be divided by GCD(a, b) $(= \text{GCD}(a, b, c))$, so that the coefficients in the reduced equation become relatively prime, or to use ancient terminology, [mutually] *dṛḍha* (firm or reduced), *niccheda* (having no divisor), *nirapavarta* (irreducible). The first example $221y + 65 = 195x$ in the *kuṭṭaka* section of the *Bījagaṇita* of Bhāskara II ([Ab, p 20]; [P, p 57]; also [PNS, p 171]) is meant to illustrate this reduction. The equation $221y - 195x = -65$ reduces to $17y - 15x = -5$. It may be pointed out to students that, by this reduction, not only do the coefficients become smaller (thereby simplifying computations), but, more importantly, one is now better equipped to tackle the problem as one has the advantage of the additional property of a and b being coprime — a property which is potentially useful.

Another subtle reduction of Āryabhaṭa II, applicable in case there is a common factor between a and c or between b and c, further reduces the size of the coefficients of x and y. Let $g_1 = \text{GCD}(a, c)$, $a_1 = \frac{a}{g_1}$, $g_2 = \text{GCD}\left(b, \frac{c}{g_1}\right)$ and $b_1 = \frac{b}{g_2}$. Then Āryabhaṭa II, and his successors like Bhāskara II, observe that the problem of solving $ay - bx = \pm c$ reduces to the problem of solving $a_1 Y - b_1 X = \pm 1$: if (u, v) is an integral solution of the latter, then $\left(\frac{cu}{g_2}, \frac{cv}{g_1}\right)$ is an integral solution of the original equation. For instance, the problem of solving $100y - 63x = \pm 90$ (cited by Bhāskara II) reduces to solving $10Y - 7X = \pm 1$. Once one obtains the solution $X = 3$, $Y = 2$ of the reduced equation $10Y - 7X = -1$ by the main algorithm (in this case, one can also get it easily by inspection!), one immediately gets the solution $x = 30$, $y = 18$ for the original equation $100y - 63x = -90$.

Second, the problem of finding all positive integral solutions is reduced to that of finding one positive integral solution. Suppose that (u, v) is a positive integral solution of $ay - bx = \pm c$. From (u, v), one first finds the minimum positive integral

---

[38] It is now a standard trick in modern mathematics to be on the lookout for simplifications or "reductions" by which efforts are made to transfer a problem to an equivalent but neater and possibly more tractable problem where the underlying features become more transparent.

solution (α, β).[39] Dividing u and v by a and b respectively, we have $u = pa + r$ and $v = qb + s$ for some whole numbers p, q, r, s such that $0 < r < a$ and $0 < s < b$. If $p = q$, then (r, s) is clearly a solution of $ay - bx = \pm c$; in fact, it is the minimum positive integral solution (proof is easy). If $p \neq q$, then it can be seen that $p < q$ when we consider the equation $ay - bx = c$; and $p > q$ when we deal with $ay - bx = -c$. The minimum positive integral solution in the two cases are $(r, s + (q - p)b)$ and $(r + (p - q)a, s)$ respectively. The above rule for arriving at the minimum solution has been explained lucidly by Āryabhaṭa II but is already implicit in Bhāskara I. If (α, β) is a minimum positive integral solution of $ay - bx = \pm c$, then Bhāskara I and his successors describe the general positive integral solution as $(\alpha + ta, \beta + tb)$ where t is a positive integer.

Third, it is clearly enough to solve an equation of the type $ay - bx = \pm 1$; for, if $av - bu = \pm 1$, then $a(cv) - b(cu) = \pm c$. Such equations were called *sthira-kuṭṭaka* (constant pulverizer). This simplification too was made by some of the Indian mathematicians right from Bhāskara I. As we shall see towards the end of this section, in problems of astronomy involving the equations $ay - bx = \pm c$, the conditions were often such that the coefficients a, b would be the same in several equations but the interpolator c would vary. In such situations, working first with the constant pulverizer and then modifying the solution according to the specific problem would have been convenient.

Some of the Indian authors also showed that we can reduce to the case $a > b$. This step was achieved through two different methods. The later method corresponds to the modern approach: if $b > a$, think of the equation as $bx - ay = \mp c$ rather than $ay - bx = \pm c$. But earlier writers like Bhāskara I, who already had to arrange the equation $ay - bx = c$ so as to have only positive coefficients, used the following device when $b > a$: Let $b = aq + b_1$ where $b_1 < a$. Then the original equation transforms into the equation $ay_1 - b_1x = c$ (where $y_1 = y - qx$) which is of the desired form. If (u, v) is a solution of $ay_1 - b_1x = c$, then $(u, v + qu)$ is a solution of $ay - bx = c$. We shall henceforth assume $a > b$ as it would facilitate our discussions.

Main Steps

We now describe the crux of Āryabhaṭa's algorithm using the interpretation of Bhāskara I. By successive division, we have $a = a_1b + r_1,\ b = a_2r_1 + r_2,\ r_1 = a_3r_2 + r_3, \cdots$ and so on. Let n denote the number of steps after which the process terminates. Since the GCD of a and b is 1, the final relation is $r_{n-2} = a_nr_{n-1} + 1$ $(1 < r_{n-1} < r_{n-2} < \cdots < r_1 < b)$. Thus, $r_n = 1$ and $r_{n+1} = 0$. Given a and b, the

[39] Since a and b are positive, if $(x_1, y_1), (x_2, y_2)$ are pairs of positive integers satisfying $ay - bx = \pm c$, then $x_1 \leq x_2$ if and only if $y_1 \leq y_2$. Therefore, one can talk about a "minimum positive integral solution" (α, β).

quantities $a_1, \cdots, a_n$ can be quickly determined by the method of "long division" for computing the GCD of a and b.

Now, for solving $ay - bx = 1$, define quantities $x_{n+2}, x_{n+1}, x_n, \cdots$ by backward induction as follows: define x_{n+2} and x_{n+1} to be whole numbers satisfying the relation $r_{n-1}x_{n+2} - x_{n+1} = (-1)^n$. Thus, if n is odd, one can simply take $x_{n+2} = 0$ and $x_{n+1} = 1$; if n is even, one can take $x_{n+2} = 1$ and $x_{n+1} = r_{n-1} - 1$. Now define x_m $(n \geq m \geq 1)$ by $x_m = a_m x_{m+1} + x_{m+2}$. For facilitating the computation of $x_n, \cdots, x_2, x_1$, the Indians constructed convenient tables called *vallī*. As will be clear from subsequent discussions, the numbers x_1, x_2 satisfy $ax_2 - bx_1 = 1$; thus (x_1, x_2) is a solution of the equation $ay - bx = 1$.

As a simple example, consider $10y - 7x = 1$ (the reduced form of one of the exercises set by Bhāskara II). Here

$$a = 10, b = 7; \text{ so that } a_1 = 1, r_1 = 3; a_2 = 2, r_2 = 1.$$

The reverse algorithm yields

$$x_4 = 1; \; x_3 = 3 - 1 = 2; \; x_2(= a_2x_3 + x_4) = 5; \; x_1(= a_1x_2 + x_3) = 7.$$

Thus we obtain the solution $x = 7$, $y = 5$.

Bhāskara I explains that one need not continue the repeated division till the stage $r_n = 1$. One can terminate at an intermediate stage k if, having obtained $r_1, \cdots, r_k$, one can observe, by inspection, positive integers u, v satisfying $r_{k-1}u - r_k v = (-1)^k$. One can then define $x_{k+2}(= u), x_{k+1}(= v), \cdots, x_2 (= y), x_1 (= x)$ recursively, as before, and obtain a solution of the equation $ay - bx = 1$.

For solving $ay - bx = -1$, one considers numbers x_{k+1} $(k \leq n)$ and x_{k+2} satisfying $r_{k-1}x_{k+2} - r_k x_{k+1} = (-1)^{k+1}$; the rest is as above. Or, one could take the approach of Bhāskara II: solve $ay - bx = 1$ and use the fact that if (α, β) is the minimum positive integral solution of $ay - bx = 1$, then $(a - \alpha, b - \beta)$ is the minimum positive integral solution of $ay - bx = -1$.

For instance, consider $10y - 7x = -1$. While $a_1, r_1, a_2, r_2(= 1)$ are as before, now the reverse algorithm yields

$$x_4 = 0, \; x_3 = 1 \text{ so that } x_2(= a_2x_3 + x_4) = 2; \; x_1(= a_1x_2 + x_3) = 3;$$

and we obtain the solution $x = 3, y = 2$. Alternatively, having computed the solution $(7, 5)$ of the equation $10y - 7x = 1$, one can immediately deduce the solution $x = 10 - 7 = 3$ and $y = 7 - 5 = 2$ for the equation $10y - 7x = -1$. Brahmagupta, Bhāskara II and Nārāyaṇa give similar rules for deriving integer solutions of $ay + bx = c$.

In section 4, we had mentioned about the efficiency of the Euclidean algorithm for determination of the GCD. As can be seen from the above steps, the extension, by *kuṭṭaka*, to the solution of the linear indeterminate equation $ay - bx = c$, is again an efficient algorithm.

The Descent-Like Idea

As indicated by Weil in [We2, p 93], there is a descent-like idea in the *kuṭṭaka* method for solving $ay - bx = \pm 1$. We may formulate it as follows (to fix our ideas, let us take the RHS to be 1):

(1) Assume the existence of a positive integral solution (x, y) to the equation $aY - bX = 1$.

(2) Then transform this equation in successive steps into equations with smaller and smaller coefficients and solutions to eventually arrive at an equation $a'Y' - X' = \pm 1$ with an obvious solution $(x', y') = (a't \mp 1, t)$ for any t.

(3) Then work backwards from this obvious solution (x', y') of the reduced equation to determine the desired solution (x, y) of the original equation.

To elaborate, assume that (x, y) is a positive solution of the equation $aY - bX = 1$. Recall that we are assuming $0 < b < a$ and a, b are coprime. If $a - b = 1$, then $(1, 1)$ is a solution of $aY - bX = 1$. So we may further assume that $a - b > 1$. Then $x > y$. If b had been 1, we would have been through. This prompts the approach (instructive for students): why not try to "break" (i.e., "pulverize") the coefficients a, b into smaller ones? Recall the relation $a = a_1 b + r_1$ with $0 < r_1 < b$. Each component of the pair (b, r_1) is smaller than that of the pair (a, b). Now one tries to transform the equation $aY - bX = 1$ to get an equation with coefficients b and r_1. The relation $ay - bx = 1$ leads to the relation $(a_1 y - x)b + r_1 y = 1$.

Set $z := x - a_1 y$. Then clearly $0 < z < y$ and (y, z) is a positive solution of the equation $bZ - r_1 Y = -1$ which is again a linear equation of the original form; but now the solution (y, z) is co-ordinate-wise smaller than the original (x, y) and, moreover, the coefficients (of the new equation) too have become smaller since $0 < b < a$ and $0 < r_1 < b$. The two pairs of solutions being linearly related, if one can determine the smaller pair (y, z), one can easily compute the original (x, y).

Denote x, y, z by x_1, x_2, x_3 respectively. Proceeding as above (introducing $x_4, x_5, \cdots$, etc.), since $r_n = 1$ (as per our earlier notation), one eventually arrives at the easily solvable equation $r_{n-1} X_{n+2} - X_{n+1} = (-1)^n$. From this equation, one works backwards to arrive at the solution (x, y) of the original equation $aY - bX = 1$.

For instance, in the example $10x_2 - 7x_1 = 1$, the *kuṭṭaka* uses the fact that solving $10x_2 - 7x_1 = 1$ is equivalent to solving $7x_3 - 3x_2 = -1$ (where $x_3 = x_1 - x_2$), which, in turn, is equivalent to solving $3x_4 - x_3 = 1$ (where $x_4 = x_2 - 2x_3$); and the latter has the obvious positive integral solution $x_3 = 2, x_4 = 1$.

Thus, the main algorithm itself is an illustration of a non-trivial application of the modern "reduction" principle: it transfers, through a sequence of steps, a somewhat involved problem to a problem with an obvious solution. Since the process involves the breaking up of the original data (both solutions and coefficients) into successively

smaller numbers by repeated division, the term *kuṭṭaka* (pulverizer) is apt; in fact, it gives a hint to the learner of the main idea behind the algorithm. Thus the *kuṭṭaka* (both the method as well as the term itself) resembles Fermat's celebrated principle of descent. As Weil remarks ([We2, p 7]):

> In later Sanskrit texts this became known as the *kuṭṭaka* (= "pulverizer") method; a fitting name, recalling to our mind Fermat's "infinite descent".

Euclidean Algorithm and *Kuṭṭaka*

There is a tendency among writers on history of mathematics to make remarks to the effect that the solution to the linear Diophantine equation "is essentially identical with" or "does not differ substantially from" or "is just a straightforward application of" the "Euclidean algorithm" for finding the GCD, and thereby to undermine (often inadvertently) the achievement involved in the discovery of the *kuṭṭaka*. As we have seen, there are two major steps or principles involved in the ancient Indian *kuṭṭaka*, as also in Bachet's solution of the linear indeterminate equation $ay - bx = c$.

I Computation of the GCD d of a and b.

II Expression of d as a linear combination of a and b.

In modern college-algebra, the two steps may be viewed, respectively, as corresponding to the algorithmic proofs of the following generalised abstract statements: (I) "$\mathbb{Z}$ is a Euclidean Domain and in a Euclidean Domain, any two non-zero elements have a GCD." and (II) "Any Euclidean Domain is a PID."[40]

Step II does not occur in the *Elements* or in any extant ancient Greek text. Its explicit presence can be seen in ancient Indian texts; in Europe, it occurs for the first time in Bachet's work (1624). Modern mathematicians, well-trained in years

---

[40]A Euclidean Domain is an integral domain R in which the "Division Algorithm" holds for any two elements a and b of R with $b \neq 0$; that is, there exists a (non-negative) integer-valued function N on R with $N(0) = 0$ such that given $a, b(\neq 0) \in R$, there exist $q, r \in R$ such that $a = bq + r$ with either $r = 0$ or $N(r) < N(b)$. A PID (Principal Ideal Domain) is an integral domain in which every ideal is principal, that is, every ideal is of the form Rd for some $d \in R$.

The concept of GCD of two integers has the following generalisation in the case of an integral domain R:- d is said to be a GCD of two non-zero elements a and b of R if (i) $a = a_1 d$ and $b = b_1 d$ for some $a_1, b_1 \in R$, and (ii) if $d' \in R$ is such that $a = a'd'$ and $b = b'd'$ for some $a', b' \in R$, then $d = d_1 d'$ for some $d_1 \in R$. As in Euclid's *Elements* (Proposition 2 of Book VII), the repeated application of Division Algorithm shows that any two non-zero elements of a Euclidean Domain have a GCD and that it can be computed algorithmically. By describing, again algorithmically, how to express the GCD as a linear combination of a and b, Āryabhaṭa's *kuṭṭaka* shows that the ideal generated by two non-zero elements a and b of a Euclidean Domain R is the principal ideal generated by their GCD thereby giving a constructive proof of the property that any Euclidean Domain is a PID (the converse is not true). For further details on Euclidean Domains and PIDs, see chapter 8 of *Abstract Algebra* by D.S. Dummit and R.M. Foote (Wiley 2003).

of high-school classical algebra, sometimes think of Step II as a trivial corollary to Step I. But to appreciate Step II in the context of its history in India, one should note that it was discovered before algebra had formally established itself in Indian mathematics and that at the time of discovery and for a few centuries thereafter, it was considered a significant feat by some of the strong ancient Indian mathematicians. Even Brahmagupta's treatise imparts considerable emphasis on the *kuṭṭaka* and provides plenty of illustrative examples. All of them touch Step I lightly, they dwell mainly on Step II. One has also to note that, after Bachet's rediscovery (1612) of the *kuṭṭaka*, Legendre had described the method as *very ingenious* as late as in 1798.[41]

A profound aspect of the *kuṭṭaka* method for solving Step II is that it contains the seeds of the idea of "descent". In fact, Bachet's rediscovery of the method might have, in a subtle way, influenced the discovery of the "descent" principle. While analyzing Brouncker's solution to the equation $Dx^2 + 1 = y^2$, Weil brings out the implicit "descent"-like idea in Brouncker's method and remarks that Bachet's method must have provided a model for Fermat and Brouncker. We quote below excerpts from the passage of Weil ([We2, p 93]); we have highlighted a portion in italics.

> The starting point here is to assume the existence of a solution (x, y) of $x^2 - Ny^2 = 1$, and to transform the problem by successive steps into others, each of which has a solution in smaller numbers; ... this is typical of Fermat's descent; the difference lies in the fact that in the latter case one seeks to prove that there is no solution, while in the former the purpose is to find one. At the same time, *Bachet's method for the solution of the equation $ay - bx = \pm 1$ undoubtedly provided a model for Wallis, Brouncker and Fermat to follow*, ...

Weil then revisits Bachet's solution of the linear indeterminate equation, bringing out its descent-like aspect. It may be noted that the discovery of the Indian *cakravāla* (which also has a descent-like aspect to it) had taken place in an environment where the *kuṭṭaka* method had a strong influence; in fact, the *kuṭṭaka* algorithm itself was used in the solution (as we shall see in section 7).

The above considerations suggest that the *kuṭṭaka* is perhaps more subtle than what its description as being "essentially the Euclidean algorithm" would indicate. Step II is of interest in computational number theory and is a fundamental step in cryptography. The equation $d = ay - bx$, expressing the GCD d as a linear

---

41 Again, one has to consider the historical context: around Bachet's time symbolic algebra had just begun to establish itself in Europe. Even F. Viète (1540–1603), who had a pioneering role in introducing symbolic algebra in Europe, did not consider negative numbers. In fact, till the 16th century, Arab and European mathematicians had struggled even with problems involving equations of the type $ax + b = c$ (where a, b, c are positive numbers) as can be seen from the prevalence of the cumbrous "rule of false position". As D.E. Smith remarks ([Sm, p. 437]): "*To the student of today, having a good symbolism at his disposal, it seems impossible that the world should ever have been troubled by an equation like $ax + b = 0$. Such, however, was the case, ...*"

combination of integers a and b, is sometimes known as *Bézout's identity* ([JJ, p 7]); perhaps *Āryabhaṭa's identity* would be more appropriate historically.

Continued Fractions

The *kuṭṭaka* may be interpreted as a technique in the theory of continued fractions. The formulation $y = \frac{bx+c}{a}$ and method of solution have given rise to a suggestion that the pioneers of the *kuṭṭaka* algorithm knew the basic principles of continued fractions ([Bg]). Knowledge of continued fractions is more apparent in some of the later Indian works. In the versions described above, after obtaining the quotients $a_1, \cdots, a_n$, one computes quantities $x_n, x_{n-1}, \cdots$ in the backward direction (i.e., one moves from below to top in the table). But in certain later texts like the *Karaṇapaddhati* and the *Yuktibhāṣā* (1540 CE) of Jyeṣṭhadeva, having obtained the quotients $a_1, \cdots, a_n$, one constructs sequences of numbers p_m and q_m by the recurrence relations

$$p_0 = 1,\ p_1 = a_1;\ p_m = a_m p_{m-1} + p_{m-2}$$

$$q_0 = 0, q_1 = 1;\ q_m = a_m q_{m-1} + q_{m-2}$$

till one reaches p_n and q_n (where n is as before). Then $aq_n - bp_n = (-1)^{n+1}$, i.e., (p_n, q_n) is a solution of one of the equations $ay - bx = \pm 1$ (the solutions of the other equation can be derived using techniques discussed earlier). Note that the quantities $a_1, a_2, \cdots, a_n, a_{n+1} (= r_{n-1})$ are the "quotients" in the continued fraction expansion of $\frac{a}{b}$; and $\frac{p_m}{q_m}$ are the successive "convergents" (which are actually characterised by the above recurrence relations). The solution of the linear Diophantine equation in the notation of continued fractions was given by Saunderson in England in 1740 and Lagrange in France in 1767.

Interpretation of Sathaye

We have discussed the traditional Indian approaches to *kuṭṭaka* due to Bhāskara I and later mathematicians. Avinash Sathaye has given an interesting and efficient version of Āryabhaṭa's algorithm that is lucidly explained, with an example, on his website ([Sat, chapter 3]; especially section 3.2). Sathaye's version can be presented as follows.

We are to find x and y such that $ay - bx = d$, where d is the GCD of a and b. Let $a_1, a_2, \cdots$ be the successive quotients and $r_1, r_2, \cdots$ be the successive remainders in the standard division algorithm for computing the GCD of a and b. Thus, if $a > b$, we have, $r_1 = a - a_1 b$, $r_2 = b - a_2 r_1 = -a_2 a + (1 + a_1 a_2) b$, and so on. Now set

$$u_0 := 0,\ v_0 := 1,\ u_1 := 1,\ v_1 := -a_1.$$

Thus $au_1 + bv_1 = r_1$. Define u_n, v_n inductively by:

$$u_n = u_{n-2} - a_n u_{n-1};\ v_n = v_{n-2} - a_n v_{n-1}.$$

Then one can see that $r_m = au_m + bv_m$ for each m; in particular, if d is reached at stage n (i.e., if $r_n = d$), then $(-v_n, u_n)$ is a solution of $ay - bx = d$. In fact, at the next stage, we get $au_{n+1} + bv_{n+1} = 0$ (since $r_{n+1} = 0$) so that $(-v_n - tv_{n+1}, u_n + tu_{n+1})$ is a solution for any integer t.

While the above algorithm fits the terse verses of Āryabhaṭa, it requires the use of negative numbers (otherwise, one has to make a careful adjustment of sign as before). One is not sure if the use of negative numbers had become sufficiently well-established (even among any elite section) during his time. But irrespective of whether Āryabhaṭa actually meant it, the above version would be a useful supplement for pedagogic purposes.

Saṁśliṣṭa Kuṭṭaka

Ancient Indian algebraists also solved simultaneous linear indeterminate equations of the type

$$b_1 y - a_1 x_1 = c_1,\ b_2 y - a_2 x_2 = c_2, \cdots, b_r y - a_r x_r = c_r$$

under the obvious consistency conditions. Note that the special case $b_1 = b_2 = \cdots b_r = 1$ is the "Chinese Remainder" problem of finding an integer y which when divided by a_1 leaves a remainder c_1, when divided by a_2 leaves a remainder c_2, and so on. In fact, Āryabhaṭa's verses actually mention the problem of simultaneous remainders (for $r = 2$). It is not clear from his brief verses whether Āryabhaṭa considered only the case $b_1 = b_2 = 1$ or the more general case; the commentary on Āryabhaṭa's verses by Parameśvara (c. 1430 CE) explains the general case itself. Bhāskara I and Brahmagupta had discussed examples where $b_1 = \cdots = b_r = 1$. This system was called *saṁśliṣṭa kuṭṭaka* (conjunct pulverizer). The generalised conjunct pulverizer was considered by later mathematicians like Mahāvīra (850 CE) and Śrīpati (1039 CE).

Bhāskara I used the term *niragra-kuṭṭākāra* (non-residual pulverizer) for the problem of solving simultaneous linear indeterminate equations; and the term *sāgra-kuṭṭākāra* (residual pulverizer) for the equivalent problem of solving simultaneous remainders. These problems have applications in computations involving periodic events — for instance, in calendar-making and in the determination of the time when a certain alignment of celestial bodies would occur (e.g., the determination of the time of eclipses). To see the connection, note that if k events $E_1, \cdots E_k$ occur regularly with periods $a_1, \cdots, a_k$ with E_i happening at times $c_i, c_i + a_i, \cdots$, then the k events occur simultaneously at time N where N is a number which, when divided

by each a_i, $(1 \leq i \leq k)$, leaves remainder c_i. An astronomy problem based on a linear indeterminate equation was called *graha-kuṭṭākāra* (planetary pulverizer).

An Application in Astronomy

Ancient Indian astronomers conceived of long time-periods (*yuga*, *kalpa*, etc) in which the five visible planets along with the Sun and the Moon were estimated to execute *integral* numbers of revolutions starting from a time when the above celestial objects were "in conjunction" — that is, all were on the great circle of the "celestial sphere" passing through the "vernal equinox" and the pole of the "ecliptic".[42] For instance, by the estimates in *Āryabhaṭīya* ([ShuS, p 6]), a *yuga* of 4320000 "sidereal years" has 1577917500 civil days and, during this period, Saturn and Mars perform, respectively, 146564 and 2296824 revolutions. Ancient astronomers must have used the *kuṭṭaka* for estimating the moments when the visible planets (and the Sun and the Moon) would be in conjunction. It is perhaps significant that Āryabhaṭa, who begins his treatise by mentioning the time of commencement of the current *yuga* and the number of revolutions performed by each planet in a *yuga*, ends his mathematics chapter by describing the *kuṭṭaka*.

We end this section by giving one example of a typical application of the *kuṭṭaka* in Indian astronomy texts. Let D denote the number of civil days in a certain epoch and N the number of revolutions performed by a planet in that epoch. Let y denote the *ahargaṇa*, the number of mean civil days elapsed since the beginning of the specified epoch up to a given day; and let x denote the *bhagaṇa*, the number of complete revolutions performed by a planet during this period. Note that x is the quotient obtained by dividing the product yN by D. The *kuṭṭaka* was applied to determine the *ahargaṇa* y and the *bhagaṇa* x from the residue $r = yN - xD$ (D and N are known constants). For instance, in *Laghu-Bhāskarīya* (chapter 8, verse 17), Bhāskara I gives certain algebraic conditions satisfied by the residues of Saturn and Mars (on a given day) and asks the reader to calculate the *ahargaṇa* (for that day) and the *bhagaṇa* of Saturn and Mars ([Shu3, p 99]). Since Saturn performs 146564 complete revolutions in 1577917500 days, on dividing the two integers by the common factor 4, we get an intermediary epoch of 394479375 days in which Saturn performs 36641 revolutions. Taking $D = 394479375$ and $N = 36641$, one gets the *sthira-kuṭṭaka* $36641Y - 394479375X = 1$ from which the solutions for y and x can be obtained depending on r. In the case of Mars, one gets the *sthira-kuṭṭaka* $191402Y - 131493125X = 1$.

[42] An explanation of the above terms is given in [Du2, Part 1: The Khagola (The Celestial Sphere) and Part 2: The Nakṣatra Dina (Sidereal Day)] and the significance of *yuga* in Indian astronomy has been discussed in [Du2, Part 3: A Brief History].

6. Bhāvanā

The Main Principle

Brahmagupta's presentation of results on quadratic indeterminate equations resembles a typical modern arrangement where one first develops the fundamental theory or principle that emerges out of the investigations into a problem, and then records the solution to the original problem(s) as one application of the basic theory. Thus, after announcing *Atha Varga Prakṛtiḥ*[43], Brahmagupta immediately states the following result ([Sha, Ch 18; verses 64–65]) which is the cornerstone of the entire section.

Theorem 1 (Brahmagupta's Bhāvanā).
If $y_i{}^2 = Dx_i{}^2 + m_i,\ i = 1, 2,$ *then* $y = Dx_1x_2 \pm y_1y_2,\ x = x_1y_2 \pm x_2y_1,\ m = m_1m_2$ *satisfy* $y^2 = Dx^2 + m.$[44]

The Appendix quotes the portion of Brahmagupta's verses pertaining to the above result. Theorem 1 is also stated in verse form by Ācārya Jayadeva, Bhāskara II (1150), Nārāyaṇa (1350), Jñanarāja (1503) and Kamalākara (1658); a proof is described in the *Bījapallava* (1548) of Kṛṣṇa ([DS2, p 148–149]).

Let D denote a fixed positive integer. For convenience, we shall use the notation $(p, q; m)$ to denote a triple of numbers satisfying $Dp^2 + m = q^2$. Then, in modern language, Theorem 1 states that the solution space of the equation $Dx^2 + m = y^2$ admits the binary operations

$$(x_1, y_1; m_1) \odot (x_2, y_2; m_2) = (x_1y_2 + x_2y_1,\ Dx_1x_2 + y_1y_2; m_1m_2).$$

$$(x_1, y_1; m_1) \star (x_2, y_2; m_2) = (x_1y_2 - x_2y_1,\ Dx_1x_2 - y_1y_2; m_1m_2).$$

Ancient Indian algebraists realised the importance of the composition laws (denoted above by $\odot$ and $\star$) and used a special technical term *bhāvanā* (composition). Jayadeva mentions the terms *samāsa-bhāvanā* (additive composition) and *viśeṣa-bhāvanā* (subtractive composition) for $\odot$ and $\star$ respectively ([Shu1, p 5]).

---

[43] *atha*: "[to commence] now". The word has the nuance of an auspicious inception which is difficult to convey in English.

[44] While the verses clearly refer to $y = Dx_1x_2 + y_1y_2$ and $x = x_1y_2 + x_2y_1$, they also appear to include $y = Dx_1x_2 \sim y_1y_2$ and $x = x_1y_2 \sim x_2y_1$. As discussed in ([Du1, p 108–109]), the expressions *saha* and *aikyaṁ* indicate "combining" Dx_1x_2 with y_1y_2 and x_1y_2 with x_2y_1. The sense of "combining" certainly includes the implication "adding"; thus the reference to the more important additive principle (producing new roots $Dx_1x_2 + y_1y_2$ and $x_1y_2 + x_2y_1$) is obvious. Presumably, "combining" also includes "taking difference" (that is, alludes to the roots $Dx_1x_2 \sim y_1y_2$ and $x_1y_2 \sim x_2y_1$). In any case, subsequent expositors like Jayadeva and Bhāskara II clearly describe the two principles successively in separate verses.

In the commentary on *Brāhma Sphuṭa Siddhānta* by Pṛthūdakasvāmin (c. 860 CE), it is repeatedly emphasised, during the exposition of Brahmagupta's verses stating Theorem 1, that the new roots (x, y, m) have been obtained from the given roots (x_1, y_1, m_1), (x_2, y_2, m_2) by *composition* ([Co1, p 363–364, footnotes]). Just before stating the *bhāvanā* formulae, Jayadeva mentions ([Sh], p 5) that *the bhāvanā operations pervade endless algorithms.*[45] The importance attached to the *bhāvanā* also comes out in the version of Theorem 1 due to Bhāskara II (see [Du1, p 111–112]) from which we quote an excerpt:

> Set down successively the lesser root $[x_1]$, the greater root $[y_1]$ and the interpolator $[m_1]$. Place under them, the same or another [triple x_2, y_2, m_2], in the same order. From them, by repeated applications of the *bhāvanā*, numerous roots can be sought. Therefore, the *bhāvanā* is being expounded. . . .

The perception of Theorem 1 as a law of composition (i.e., in terms of an operation like $\odot$) comes out in the applications by Brahmagupta and others, in the version of Theorem 1 due to Bhāskara II and, above all, in the choice and use of the term *bhāvanā*.

In general, the Sanskrit word *bhāvanā* has several meanings including "production" and "demonstration"; in the context of mathematics, *bhāvanā* means "composition" or "combination". A similar Sanskrit word *bhāvita* (usual meaning: "created", "produced"; or "future", "to be") was adopted in Indian algebra by Brahmagupta and others to denote the *product* of different unknown quantities. The law $\odot$ (or $\star$) was perceived as a special mathematical *operation* involving a special *multiplicative* principle. In fact, Udaydivākara (1073 CE) explains *bhāvanā* as multiplication ([Shu1, p 6]).

While the precise sense of *bhāvanā*, in the context of $\odot$ (or $\star$), is that of a principle of "composition" — a sort-of generalised "product" — the term also carries the additional suggestion of being a principle of "production" (i.e., "generation") of new roots, and of being a powerful "lemma" (i.e., "producer" of new results). The description of Theorem 1 by Brahmagupta and other Indian algebraists are invariably followed by results crucially involving applications of the *bhāvanā* principle $\odot$. In view of the importance of the laws like $\odot$, an appropriate terminology was chosen which would be rich in significance — simultaneously conveying the sense of a multiplicative principle of composition, a principle of generation, a special lemma, etc.

Theorem 1 can be formulated in the form of the identities

$$({y_1}^2 - D{x_1}^2)({y_2}^2 - D{x_2}^2) = (Dx_1x_2 \pm y_1y_2)^2 - D(x_1y_2 \pm x_2y_1)^2$$

which are sometimes called Brahmagupta's identities.

---

[45] In the light of modern developments, Jayadeva's statement has turned out to be prophetic.

Some Applications of the *Bhāvanā* by Brahmagupta

After stating Theorem 1, Brahmagupta gives useful offshoots. First, he indicates ([Sha, Ch 18, verses 64–65]) how to construct infinitely many rational solutions of $Dx^2 + 1 = y^2$ (to be discussed in next section). Next he points out ([Sha, Ch 18, verse 66]) that infinitely many integer solutions of $Dx^2 + m = y^2$ can be generated from a given solution (if it exists) and a solution of $Dx^2 + 1 = y^2$ (see below for examples with $D = 13$); in particular, from one integer solution of $Dx^2 + 1 = y^2$, one can generate infinitely many integer solutions ([Du1, p 90–93]). And then, Brahmagupta gives his partial solution to the original problem ([Sha, Ch 18, verses 67–68]; [Du1, p 93–95]).

Theorem 2 (Brahmagupta).

(i) If $Dp^2 + 4 = q^2$, then $\left(\frac{1}{2}p(q^2 - 1), \frac{1}{2}q(q^2 - 3)\right)$ is a solution of $Dx^2 + 1 = y^2$.

(ii) If $Dp^2 - 4 = q^2$, and $r = \frac{1}{2}(q^2 + 3)(q^2 + 1)$, then $(pqr, (q^2 + 2)(r - 1))$ is a solution of $Dx^2 + 1 = y^2$.

Let $(p, q; m)^n$ denote $(p, q; m) \odot (p, q; m) \cdots \odot (p, q; m)$ n times. The results (i) and (ii) can be obtained by considering the products $(p, q; 4)^3$ and $(p, q; 4)^6$ respectively and making necessary substitutions and simplifications. For instance, for (i), using the relations $(p, q; 4) \odot (p, q; 4) = (2pq, Dp^2 + q^2; 4^2)$ and $Dp^2 = q^2 - 4$, and cancelling the common factor 4 from the consequent identity $D(2pq)^2 + 4^2 = (2q^2 - 4)^2$, one gets the triple $(pq, q^2 - 2; 4)$. Next, considering the composition $(p, q; 4) \odot (pq, q^2 - 2; 4)$ and making adjustments as above, one would arrive at the triple $\left(\frac{1}{2}p(q^2 - 1), \frac{1}{2}q(q^2 - 3); 1\right)$.

The results are illustrated by numerical examples which are usually in the spirit of "good exercises" — demanding a degree of ingenuity — rather than mechanical applications. We quote below one example ([Sha, Ch 18, verses 71–72]) which illustrates the technical power of Theorem 1.

Illustrative Examples

Example 1 (Brahmagupta). *Solve, in integers, the equation $92x^2 + 1 = y^2$.*

Solution. One readily observes that $92 \times 1^2 + 8 = 10^2$. By *samāsa-bhāvanā*,

$$(1, 10; 8) \odot (1, 10, 8) = (20, 192; 8^2).$$

Dividing the consequent identity $92 \times 20^2 + 8^2 = 192^2$ by 8^2, one obtains the

triple $\left(\frac{5}{2}, 24; 1\right)$. Now

$$\left(\frac{5}{2}, 24; 1\right) \odot \left(\frac{5}{2}, 24; 1\right) = (120, 1151; 1),$$

an integer triple. Thus (120, 1151) is a solution of $92x^2 + 1 = y^2$. ■

The thrill of Brahmagupta at his discovery can be felt from the phrase he used after stating the equation $92x^2 + 1 = y^2$: *kurvannāvatsarād gaṇakaḥ* — "One who can solve it within a year [is truly a] mathematician."

Note that in this example $m = 8$ in the initial triple whereas Theorem 2 is a statement for $m = \pm 4$. Through the example, Brahmagupta conveys to the reader that, by clever algebraic manipulations, Theorem 1 can be made to cover various cases where the initial triple is not necessarily of the form $(p, q; \pm 4)$.

Another remarkable example of Brahmagupta ([Sha, Ch 18, verses 71–72]) is the equation $83x^2+1 = y^2$. Applying Theorems 1 and 2 on the identity $83 \times 1^2 - 2 = 9^2$, one gets the solution $(x, y) = (9, 82)$ from which one can generate a sequence of successively larger solutions using Theorem 1; one of the solutions would be $x = 175075291425879$, $y = 1595011813884802$ (see [OR]).

Through this example, Brahmagupta expects the reader to observe that since $(\pm 2)^2 = 4$, successive applications of Theorems 1 and 2 on an integer identity $Da^2 \pm 2 = b^2$ will give an integer solution to the equation $Dx^2 + 1 = y^2$. A later mathematician Śrīpati clearly formulates ([Sih, p 40]; [DS2, p 157]) this immediate consequence of Brahmagupta's results: *From any positive integer solution of* $Dx^2 + m = y^2$, *where* $m \in \{-1, \pm 2, \pm 4\}$, *one can derive a positive integer solution of* $Dx^2 + 1 = y^2$ *by repeated use of the samāsa-bhāvanā.*[46]

As an illustration of the principle of generating larger roots of $Dx^2 + m = y^2$ from a given root, Brahmagupta mentions the equations $13x^2 + 300 = y^2$ and $13x^2 - 27 = y^2$ ([Sha, Ch 18, verses 75]). By inspection, one has solutions (10, 40) and (6, 21) respectively. Also from the relation $13 \times 1^2 - 4 = 3^2$, one gets the solution (180, 649) of $13x^2 + 1 = y^2$ by Theorem 2. Now taking the composite of (10, 40; 300) with (180, 649; 1), one gets the larger solution (13690, 49360) of $13x^2 + 300 = y^2$; further compositions will yield still larger solutions. Similarly, for $13x^2 - 27 = y^2$, one considers $(6, 21; -27) \odot (180, 649; 1)$, $(6, 21; -27) \odot (180, 649; 1) \odot (180, 649; 1)$, and so on.

In section 1, we mentioned that it is now known that all positive integer solutions of $Dx^2 + 1 = y^2$ are given by (x_n, y_n) satisfying $y_n + \sqrt{D}x_n = \left(y_1 + \sqrt{D}x_1\right)^n$, where (x_1, y_1) is the fundamental solution (i.e., the minimal positive integral solution). Once the fundamental solution is determined, one can see that the other

[46]This gives another strategy for solving Example 1: it is enough to solve $92x^2 + 4 = y^2$ and hence $23x^2 + 1 = z^2$; use $23 + 2 = 5^2$.

solutions (x_n, y_n) are given by Theorem 1. The sequence $\left\{\frac{y_n}{x_n} \mid n \geq 1\right\}$ forms a sequence of "best" rational approximations to $\sqrt{D}$ (the precise meaning will be stated in section 7).

Some Perspectives on the *Bhāvanā*

We now recall a few points that are further elaborated in the author's article [Du1].

(i) Brahmagupta's very approach to the problem anticipates a modern trend (which perhaps began with Lagrange): the quest for general principles, the attempt to view a problem as part of a larger set-up. While trying to solve a specific hard problem $Dx^2 + 1 = y^2$ (in two variables), Brahmagupta undertook a bold and farsighted exploration of the *general* picture: the solution space of $Dx^2 + m = y^2$ (in three variables). In the process Brahmagupta discovers and extracts an important general and abstract principle (Theorem 1), and makes a clear enunciation of this principle.

(ii) Brahmagupta envisages the key ingredient of a modern abstract structure: *binary composition*. If we exclude the four elementary arithmetic operations, the *bhāvanā* is perhaps the first conscious construction of a binary composition. Note that the binary operation is quite a complicated one: it involves *two* integral *triples* of *unknown roots*.[47] Thus, in an attempt to solve an indeterminate equation in two variables, a seventh-century mathematician thought of constructing, what amounts to, an intricate abstract structure on the solution space of an equation in three variables.[48] One does not see such an approach in mathematics during the next 1000 years. The discovery of an algebraic structure, such as Brahmagupta's composition law, on a set of significance, is now an important theme in modern mathematics research.

(iii) The power of the general composition principle of Brahmagupta can be seen from the ease with which it immediately provides the solution of a difficult equation like $92x^2 + 1 = y^2$. Ironically, this very simplicity tends to cloud the historian's perception of the true worth of Theorem 1. A casual observer may miss the richness and magnificence of ideas encapsulated in the 2-step solution of $92x^2 + 1 = y^2$. Further, once the mathematical community gets accustomed to an original idea like the *bhāvanā*, it becomes all the more difficult to fathom the greatness of the discovery. Historians need to be aware of this intrinsic risk of missing the real depth and significance of Brahmagupta's work on the *varga-prakṛti*.

---

[47] Recall that even basic symbolic computations with unknown *roots* — treating them as if they were known quantities — is a fairly modern approach (from late 18th century) that emerged during the investigations on the general polynomial in *one* unknown.

[48] The idea occurs in a fluid amorphous form; it had not been crystallised in a precise set-theoretic framework — the solution space was not presented as a single set.

(iv) The *bhāvanā* influenced, directly and indirectly, subsequent research on indeterminate equations by Indian algebraists. Apart from providing partial solutions to the original problem, the *samāsa-bhāvanā* contained the key to the subsequent complete solution by the *cakravāla* algorithm as we will see in the next section.

(v) Sometimes the brilliance of a particular algebraic research work lies in its opening up of new and unexpected horizons with immense possibilities through surprisingly simple innovations. Let us view the *samāsa-bhāvanā* in this light. Not only did Theorem 1 influence the complete solution to the original problem, it had opened up new possibilities whose true potential began to be harnessed only after its rediscovery in the 18th century. Theorem 1 turned out to be "a theorem of capital importance" (in the words of Euler[49]). For instance, the *samāsa-bhāvanā* $\odot$ has an elegant interpretation in terms of the norm function — a very important concept in modern mathematics. Let $A = \mathbb{Z}\left[\sqrt{D}\right] \left(= \left\{b + a\sqrt{D} | a, b \in \mathbb{Z}\right\}\right)$, where $\mathbb{Z}$ denotes the set of integers. The norm function on A is the map $N : A \to \mathbb{Z}$ defined by

$$N\left(y + x\sqrt{D}\right) = \left(y + x\sqrt{D}\right)\left(y - x\sqrt{D}\right) = y^2 - x^2 D.$$

Brahmagupta's identity may be reformulated as the statement:

The norm function N is multiplicative, i.e., $N(\alpha\beta) = N(\alpha)N(\beta)\ \forall \alpha, \beta \in A$.

Now the solution space $S := \{(x, y, m) \in \mathbb{Z}\times\mathbb{Z}\times\mathbb{Z} \mid Dx^2 + m = y^2\}$ is precisely the graph of N and we have a bijection $f : S \to A$ defined by $f(x, y, m) = y + x\sqrt{D}$. Transferring the ring multiplication on A to S via f^{-1}, we get the *samāsa-bhāvanā*. Thus the exotic multiplicative structure $\odot$ on S actually corresponds to the natural ring multiplication on A.

It is tempting to surmise that Brahmagupta discovered Theorem 1 through an algebraic manipulation which was, in essence, the verification of the structure preserving property of the norm function — the natural multiplication in A providing the precise formula (see [Du1, p 99–101]).

(vi) In the language of quadratic forms, Theorem 1 says that two binary quadratic forms with a given discriminant D (say $y^2 - Dx^2$, $v^2 - Du^2$) can be *composed*[50] to yield another such form with discriminant D in a new pair of variables ($xv \pm$

---

[49]Euler highlighted the result as *theorema eximium*; see ([We2, p 204–205]).

[50]The theory of composition of quadratic forms, an important and rich topic initiated by Gauss and Dirichlet, is still an active area of research. In recent years, Manjul Bhargava discovered higher composition laws, a work which has many important ramifications. An account is given in Bhargava's article *Higher composition laws and applications* in the Proceedings of International Congress of Mathematicians (ICM), Madrid (2006).

yu, $yv \pm Dxu$). Theorem 1 can also be viewed as the result (see [Du1, p 83–84] for details):

The binary form $y^2 - Dx^2$ is strongly multiplicative (over rationals).

This version of Theorem 1 was generalised in 1965 by A. Pfister using, what are now called, "Pfister forms". Pfister's discovery opened up new directions in the theory of quadratic forms. Theorem 1 thus happens to be a starting point of a distinct area of research in algebra. In his 1990 research monograph [Oj] in this field, Manuel Ojanguren begins Chapter 5 by quoting Brahmagupta's original Sanskrit verses describing Theorem 1; the chapter itself is titled *Also sprach Brahmagupta*[51].

(vii) While it is exciting to dwell on the implicit occurrence of modern principles in an ancient text, let us not lose sight of one aspect of the section *Varga-prakṛti*: the sheer wizardry in classical algebraic manipulation at an early stage in the history of symbolic algebra.

(viii) As argued in section 3 of our earlier paper ([Du1]), Brahmagupta's treatment of *varga-prakṛti* has a unique pedagogic potential. Theorem 1 defines an intricate binary operation $\odot$ on $S = \{(x, y, m) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \mid Dx^2 + m = y^2\}$. This sophisticated idea of constructing a binary composition on an abstractly defined unknown set is the quintessence of modern "abstract algebra". Theorem 1 and Example 1 can be effectively used to convey to a student the power of the concept "binary composition" and help him appreciate the essence of abstract algebraic ideas. It could be inspiring for the student to realise that the simplicity of the solution to Example 1 has its secret in the monoid structure of $(S, \odot)$ which in turn has its root in the natural ring structure of the set A (defined above in (v)). As we mentioned in [Du1, p 88], "the unexpected invocation of an abstract algebraic technique for a concrete problem by an ancient mathematician could infuse in him [a student] a dynamic and creative vigour in his formal study of abstract algebra. Thus, through this deep but easily accessible work of a Master, the perceptive student can get an early exposure to modern sophistication, see a natural application of abstract algebraic principles, develop a flair for exploring worthwhile generalisation, and truly imbibe the spirit of the abstract algebra culture without getting swept away by its formalism."

(ix) Due to the paucity of adequate historical materials, it is often difficult to ascertain whether a particular result in a treatise was discovered by the author or by a predecessor. One does not know who was the true discoverer of the *cakravāla*; even one cannot be sure whether Āryabhaṭa was the pioneer of *kuṭṭaka*. But the *bhāvanā* has the unmistakable stamp of Brahmagupta's genius. One does not see any seed of this idea in the treatises of his predecessors Āryabhaṭa and Bhāskara

[51] In English: Thus Spake Brahmagupta.

I or in the Bakṣālī manuscript. None of the results in these remarkable texts come close to the algebraic sophistication shown by Brahmagupta's research on *varga-prakṛti*. In several other results of Brahmagupta, one discerns a similar abrupt leap in mathematical maturity.

7. Cakravāla

In the *Laghu-Bhāskarīya* (Ch 8; verse 18) of Bhāskara I, there is a problem from astronomy involving the simultaneous quadratic indeterminate equations $8x + 1 = y^2$ and $7y^2 + 1 = z^2$ ([Shu3, p 101]). By inspection, one can quickly arrive at the integral solution $y = 3$, $z = 8$ for the second equation. In the context of this special equation, Udaydivākara, in his commentary on the *Laghu-Bhāskarīya*, refers to Jayadeva's general *cakravāla* method for solving any equation $Dx^2 + 1 = y^2$. Jayadeva's verses on the *cakravāla* algorithm have been published in ([Shu1]); a portion has been reproduced in the Appendix. Bhāskara II's verses on this algorithm occur in *Bījagaṇita* ([Ab, p 23], [P, p 76]); they have been quoted in M.S. Sriram's article [Srr, p 167].

The Algorithm

Recall that $(p, q; m)$ denotes a triple satisfying $Dp^2 + m = q^2$. The underlying idea of the algorithm can then be put as follows: from an integer triple $(p_n, q_n; m_n)$ such that $|m_n|$ is "small", one has to construct an integer triple $(p_{n+1}, q_{n+1}; m_{n+1})$ with $|m_{n+1}|$ "small" eventually arriving at an integer triple of the form $(p, q; 1)$. It will be convenient for our discussions to assume that, at each stage, the integers p_n and m_n are mutually coprime. Observe that, if, in an integer triple $(p, q; m)$, the integers p and m have a GCD g, then the relation $Dp^2 + m = q^2$ will show that g is a factor of q and g^2 a factor of m and hence $\left(\frac{p}{g}, \frac{q}{g}; \frac{m}{g^2}\right)$ will also be an integer triple. Thus, replacing $(p, q; m)$ by $\left(\frac{p}{g}, \frac{q}{g}; \frac{m}{g^2}\right)$ if necessary, we may ensure that p and m are relatively prime.

Integer triples of the type $(1, y; y^2 - D)$ were already "in the air" among Indian algebraists who expounded on Brahmagupta's work.[52] The authors of *cakravāla* invariably begin with an initial triple $(p_0, q_0; m_0)$ of this type: $p_0 = 1$, q_0 a positive integer for which $|{q_0}^2 - D|$ is minimum, and $m_0 = {q_0}^2 - D$.

---

[52]Śrīpati (and successive algebraists) clearly explains — what is tersely indicated by Brahmagupta — that applying *bhāvanā* on two copies of any triple $(1, q; q^2 - D)$, one gets the rational solution $x = \frac{2q}{q^2 \sim D}$, $y = \frac{q^2+D}{q^2 \sim D}$ of $Dx^2 + 1 = y^2$ and that infinitely many roots of this equation can be obtained by varying q and also by repeating the *bhāvanā* ([Sih, p 40]; [DS2, p 153;150]). The possible impact of this observation on the discovery of the *cakravāla* is discussed in [Du1, p 90–92; 96–97].

We now explain the inductive step. Suppose that one has completed n steps and arrived at an integer triple $(p_n, q_n; m_n)$ with p_n and m_n coprime. Composing this (known) integer triple $(p_n, q_n; m_n)$ with the (variable) integer triple $(1, y; y^2 - D)$ by *samāsa-bhāvanā*, one gets an integer triple $(p_n y + q_n, Dp_n + q_n y; m_n(y^2 - D))$, that is, the identity

$$D(p_n y + q_n)^2 + m_n(y^2 - D) = (Dp_n + q_n y)^2.$$

Dividing the above identity by $m_n{}^2$ (an idea which too could have arisen from the methods for obtaining rational solutions), one obtains a rational triple

$$(p_{n+1}, q_{n+1}; m_{n+1}) = \left(\frac{p_n y + q_n}{m_n}, \frac{Dp_n + q_n y}{m_n}; \frac{y^2 - D}{m_n}\right).$$

Now the identity $q_n{}^2 = Dp_n{}^2 + m_n$ shows that

$$m_n p_{n+1} q_n = (p_n y + q_n) q_n = p_n(Dp_n + q_n y) + m_n.$$

Suppose that p_{n+1} is an integer. As p_n and m_n are coprime, it would then follow that m_n divides $Dp_n + q_n y$, that is, q_{n+1} is an integer. It also follows that $m_{n+1}(= Dp_{n+1}^2 - q_{n+1}{}^2)$ is an integer. Further, one has the relation $p_{n+1}q_n - p_n q_{n+1} = 1$ which shows, by the argument made at the outset, that p_{n+1} and m_{n+1} are coprime.

Thus $(p_{n+1}, q_{n+1}; m_{n+1})$ would be an integer triple if y can be so chosen that p_{n+1} (and hence q_{n+1}, m_{n+1}) becomes an integer. This amounts to finding integer solutions of the equation $m_n x - p_n y = q_n$. As we saw in section 5, the linear indeterminate equation had been extensively discussed by Indian algebraists from the time of Āryabhaṭa (499 CE) and they knew the complete solution to the problem. Now among the infinitely many solutions, Jayadeva and Bhāskara II choose a solution for which $|y^2 - D|$ would be minimum so that $|m_{n+1}|$ is minimised (as desired).

It can be shown that the process terminates after a finite number of steps (that is, $m_n = 1$ for some n), but the proof does not occur in any ancient text. Note that it is enough to arrive at a stage in which $m_n \in \{\pm 1, \pm 2, \pm 4\}$; for, in that case, one can use Brahmagupta's formulae (Theorem 2) to arrive at a solution of $Dx^2 + 1 = y^2$. The ancient Indian authors assert that one will always arrive at such a stage.

Starting with the initial triple $(p_0, q_0; m_0)$, the *cakravāla* (with or without the *bhāvanā*) invariably leads one to the fundamental solution of $Dx^2 + 1 = y^2$ mentioned in section 1 ([Sel, p 177]) and, as mentioned earlier, repeated applications of Theorem 1 then yield *all* integer solutions.

An Example

Quite often Brahmagupta's result gives a very effective short-cut for numerical computations as the solution of the famous example of Bhāskara II ([Ab. p 24]; [P, p 77]) shows:

Example 2 (Bhāskara II). *Solve, in positive integers,* $61x^2 + 1 = y^2$.

Solution. As 64 is the perfect square nearest to 61, we have the initial triple $(1, 8; 3)$. Now one finds positive integer y for which $\frac{y+8}{3}$ is an integer and $|y^2 - 61|$ is minimised. Clearly $y = 7$. Now

$$\frac{7+8}{3} = 5; \ \frac{61 + 8 \times 7}{3} = 39; \ \frac{7^2 - 61}{3} = -4.$$

Thus we have the second triple $(5, 39; -4)$. Now rather than continuing the *cakravāla*, one can apply Brahmagupta's formula (Theorem 2) on the triple $(5, 39; -4)$ to immediately get the minimum positive integral solution (226153980, 1766319049) of $61x^2 + 1 = y^2$. ■

The article ([OR]) works out solutions to the examples of Nārāyaṇa using the *cakravāla* aided by the *bhāvanā*. The minimum positive integral solution for $103x^2 + 1 = y^2$ is (22419, 227528) and for $97x^2 + 1 = y^2$ is (6377352, 62809633).

Application to Rational Approximation

Nārāyaṇa applies the *cakravāla-bhāvanā* method of generating arbitrarily large solutions of $Dx^2 + 1 = y^2$ to determine progressively better rational approximations of $\sqrt{D}$ ([D1]; [Du1, p 93; 102-3]). Note that, if $Da^2 + 1 = b^2$, then

$$\frac{b}{a} - \sqrt{D} = \frac{b^2 - Da^2}{a(b + a\sqrt{D})} = \frac{1}{a(b + a\sqrt{D})}$$

and thus, for a sufficiently large solution (a, b), $\frac{b}{a}$ will be a good approximation for $\sqrt{D}$. Nārāyaṇa cites two numerical examples: $\sqrt{10}$ and $\sqrt{\frac{1}{5}}$. We discuss the former; the case of $\sqrt{\frac{1}{5}}$ is similar.

For $\sqrt{10}$, Nārāyaṇa mentions the rational approximations $\frac{19}{6}$, $\frac{721}{228}$ and $\frac{27379}{8658}$. To see how the fractions arise, consider the notation of section 6 for $D = 10$. Since 9 is the perfect square nearest to 10, one has the initial triple $(1, 3; -1)$. Now,

$$(1, 3; -1) \odot (1, 3; -1) = (6, 19; 1),$$

$$(6, 19; 1) \odot (6, 19; 1) = (228, 721; 1),$$

$$(6, 19; 1) \odot (228, 721; 1) = (8658, 27379; 1).$$

Thus one has the three successive fractions $\frac{19}{6}$, $\frac{721}{228}$ and $\frac{27379}{8658}$ as rational approximations to $\sqrt{10}$. The usefulness of the method can be seen from the fact that the third fraction $\frac{27379}{8658}$ ($= 3.162277662\ldots$) matches the value of $\sqrt{10}$ ($= 3.162277660\ldots$) up to nine decimal digits. This method of generating successively closer approximations was restated by Euler in 1732 ([D1, p 188]). It is now known that if $\{(x_n, y_n) \mid n \geq 1\}$ is the sequence of positive integral solutions of $Dx^2 + 1 = y^2$, then, among all fractions with denominator x_n or less, $\frac{y_n}{x_n}$ is the best approximation to $\sqrt{D}$. This follows from a basic theorem in continued fractions ([BrC, p 399–401; 535]; [G, p 63–64; 71–72]). In particular, Nārāyaṇa's approximations for $\sqrt{10}$ are the best possible (with the respective restrictions on the denominators). One can similarly see that the *Śulba* approximations for $\sqrt{2}$, mentioned in section 1, are again the best possible. For the equation $2x^2 + 1 = y^2$, the minimum positive integral solution is clearly (2, 3). Now the *bhāvanā* relations

$$(2, 3; 1) \odot (2, 3; 1) = (12, 17; 1) \text{ and } (12, 17; 1) \odot (12, 17; 1) = (408, 577; 1)$$

yield the two *Śulba* fractions $\frac{17}{12}$ and $\frac{577}{408}$. It can also be seen, directly or from the theory of continued fractions, that the *Śulba* fraction $\frac{7}{5}$ is the best approximation to $\sqrt{2}$ among all fractions with denominator ≤ 5.

Some Perspectives on the *Cakravāla*

We now make a few observations.

(i) The *cakravāla* algorithm demonstrates a marvellous interplay between the two great preceding works — the *kuṭṭaka* and the *bhāvanā*. As Weil observes ([We2, p 22]):

> As is the case with many brilliant discoveries, this one [cakravāla] can be seen in retrospect as deriving quite naturally from the earlier work [samāsa-bhāvanā].

The pedagogic potential of this development can hardly be overstated. A student can be shown an "illustrious example"[53] of how first-rate research flows: how Āryabhaṭa seizes upon an ancient algorithm for computing the GCD to make a "descent" into the *kuṭṭaka* algorithm for solving the linear Diophantine equation, a problem with applications in astronomy; how Brahmagupta approaches the harder problem of solving the Pell equation and comes up with a brilliant *bhāvanā* yielding a partial solution to the problem, and how Jayadeva (or a predecessor) assimilates both the ideas to arrive at the astonishing *cakravāla* algorithm.

---

[53] The allusion is to a statement of Leibniz quoted by Weil ([We1, p 229]; [Du1, p 88]).

(ii) Expositions on the *cakravāla* often tend to identify it with the Brouncker-Euler-Lagrange procedure. But there is a difference. In the notation used above, the Brouncker-Euler-Lagrange method would amount to choosing quantities $y < D$ for which $D - y^2$ is minimised (subject to conditions above), while the *cakravāla* of Jayadeva and Bhāskara II stipulates that $|D - y^2|$ is minimised.[54] The steps in the former method correspond to certain convergents of the standard continued fraction expansion of $\sqrt{D}$; the steps in the latter correspond to a generalised expansion where negative remainders could be involved. For instance, for the case $D = 61 (= 7^2 + 12 = 8^2 - 3)$, the starting point in the former procedure will be 7, while in the latter it will be 8.

The *cakravāla* can also be interpreted in the framework of the regular (i.e., the usual) continued fraction expansion of $\sqrt{D}$ (see [Sig]); its steps correspond to fewer convergents of the expansion than the steps of Brouncker's method. For instance, the first three terms of the continued fraction expansion of $\sqrt{61}$ are $\frac{7}{1}$, $\frac{8}{1}$ and $\frac{39}{5}$ (with $\frac{7}{1} < \frac{39}{5} < \sqrt{61} < \frac{8}{1}$); the *cakravāla* involves only $\frac{8}{1}$ and $\frac{39}{5}$. As discussed in section 5, there are indications that ancient Indian mathematicians had possibly discovered principles of continued fractions. The term *cakravāla* (cyclic) could be an allusion to the periodicity of a continued fraction expansion of $\sqrt{D}$ (see [Sr1, p 286]) or some equivalent concept.

(iii) The descriptions of *cakravāla* by Jayadeva and Bhāskara II begin with the inductive step of constructing $(p_{n+1}, q_{n+1}; m_{n+1})$ from $(p_n, q_n; m_n)$ without mentioning any starting point $(p_0, q_0; m_0)$. Possibly the authors of *cakravāla* were flexible regarding the choice of the initial triple $(p_0, q_0; m_0)$; *any* three convenient integers p, q, m satisfying $Dp^2 + m = q^2$ could be the starting point.[55] They of course had an obvious choice for the initial triple, that is, a triple of the form $(1, q; q^2 - D)$ which had already become common in the study of *varga-prakṛti*. In the illustrative examples, they invariably display the triple $(1, q_0; {q_0}^2 - D)$ where q_0 is an integer y for which $|y^2 - D|$ is minimum. As triples of the form $(1, y; y^2 - D)$ were highlighted, in any case, as a central tool in the *cakravāla* verses, a separate mention of $(1, q_0; {q_0}^2 - D)$ was perhaps felt unnecessary.[56] Incidentally the triple $(1, q; q^2 - D)$ finds explicit mention in Jayadeva's verse on the solution of $Dx^2 + m = y^2$ occurring immediately after his verses on $Dx^2 + 1 = y^2$ ([Shu1, p 14]).

Perhaps some of the writers of books on History of Mathematics, who erroneously state that the *cakravāla* involves "trial-and-error", get confused by the

---

[54]Nārāyaṇā appears to differ slightly on this point. While he is silent on the minimisation in his description of the *cakravāla*, in his examples, he chooses y for which $|y - \sqrt{D}|$ is minimised; see [Sig].

[55]Recall that in the *kuṭṭaka* too there was a flexibility of stopping at some stage k (rather than completing the long division for computation of the GCD) if one could see, by inspection, the solution of some intermediary linear equation.

[56]Recall the culture of brevity in ancient Indian treatises discussed at the beginning of section 5.

apparent obscurity (in the translations of the verses) regarding the initial triple. They do not realize that triples $(1, q; q^2 - D)$ were taken for granted by the ancient authors of *cakravāla*.[57]

(iv) As observed by A.A. Krishnaswamy Ayyangar ([Kr1, p 234]), in retrospect, the original *cakravāla* can be simplified by avoiding the *kuṭṭaka*. Denote by y_{n+1} the integer y which satisfies $m_n p_{n+1} - p_n y_{n+1} = q_n$. Thus

$$p_n = \frac{p_{n-1} y_n + q_{n-1}}{m_{n-1}}; \; q_n = \frac{D p_{n-1} + q_{n-1} y_n}{m_{n-1}}; \; m_n = \frac{{y_n}^2 - D}{m_{n-1}}$$

so that $p_n y_n - q_n = p_{n-1} m_n$ and hence

$$p_{n+1} \left(= \frac{p_n y_{n+1} + q_n}{m_n} \right) = \frac{p_n (y_{n+1} + y_n)}{m_n} - p_{n-1};$$

in particular, $y_{n+1} + y_n \equiv 0 \pmod{m_n}$. Hence, instead of finding p_{n+1} by *kuṭṭaka*, one can construct y_{n+1} inductively as follows: start with $p_0 = 1$, $y_0 = q_0$ a positive integer y for which $|y^2 - D|$ is minimum, and $m_0 = {y_0}^2 - D$; having constructed y_n, p_n, q_n, m_n with p_n, q_n, m_n being mutually coprime, choose y_{n+1} to be a positive integer y satisfying the simple congruence $y \equiv -y_n \pmod{m_n}$ for which $|y^2 - D|$ is minimal. Then $p_{n+1} := \frac{p_n(y_{n+1}+y_n)}{m_n} - p_{n-1}$ is an integer; and, as before, the corresponding q_{n+1}, m_{n+1} will also be integers with $p_{n+1}, q_{n+1}, m_{n+1}$ being mutually coprime.

Simplifications are observed *after* initial breakthroughs. But, as mentioned earlier, there is a certain richness of thought in the ingenious approach through the blending of *kuṭṭaka* and *bhāvanā* which would be good for students to imbibe. For the purpose of promoting the art of discovery, it is desirable to highlight the initial thought-process which actually achieved the breakthrough and which tends to get camouflaged in later simplifications.

(v) As pointed out in [Kr1], the inductive step in the *cakravāla* algorithm involves a transformation from one quadratic form to an *equivalent* quadratic form.

(vi) Weil makes an incidental remark ([We2, p 51]):

> the Indians' love for large numbers may have supplied a motivation for their treatment of Pell's equation.[58]

---

[57]Another source for the confusion could be that both Jayadeva and Bhāskara II first describe Brahmagupta's *bhāvanā* method which is a statement about generating new solutions out of given solutions without saying how to get a given solution.

[58]It is significant that this observation has been made by Weil who had considerable familiarity with ancient Indian literature.

Indeed, Indians had a fascination for large numbers from the Vedic times.[59] The facility with large numbers was conducive for the Indians in their research on the Pellian equation, which soon involves large solutions.

(vii) The early occurrence of Pell's equation in Indian history can also be seen in the light of an observation of Legendre ([G, p 3]):

> . . . it appears that the ancient philosophers[60] did rather extensive research on the properties of numbers. But they lacked two tools needed to reach the depths of this science: the art of numeration, which serves to express numbers with great ease, and algebra, which generalizes results and can operate equally on the known and the unknown. The invention of both these arts greatly influenced the progress of the science of numbers.

Greeks had generally neglected computational mathematics and it is significant that Archimedes, who had suggested a problem which was later seen to involve Pell's equation, also happened to be the one Greek mathematician who had realised the importance of an efficient system of enumeration and had tried to evolve one.

While the "art of numeration" had been perfected in India well before the time of Āryabhaṭa, the Indian algebraists from Brahmagupta onwards had also become adept in the art of formation and clearance of equations. Brahmagupta used the term *samīkaraṇa* (making equal) for equation, *varṇa* (colour-names) for unknowns and *rūpa* (appearance) for coefficients, described a plan for writing equations and then preparing them, and classified equations. Further details are given in [DS2].

(viii) Perhaps it can never be ascertained whether ancient Indian algebraists had conceived of a proof[61] that the *cakravāla* method always works. Even if they had a proof, given the technicalities involved, it would have been difficult to express it in those days. After all, even Fermat had not written (or is not known to have written) a proof, although he claimed that he had a proof by "descent". While analyzing what might have been Fermat's proof, Weil provides a valuable insight into Fermat's possible difficulty in formulating it ([We2, p 99]):

> It does not seem unreasonable to assume that this may have been, in general outline, what Fermat had in mind when he spoke of his proofs in his letters to Pascal and Huygens; how much of it he could have made explicit must remain a moot question. Perhaps one of his

---

[59] A verse of Medhātithi in the *Vājasaneyī Saṁhitā* (XVII.2) of the *Śukla Yajurveda* gives number-names for each power of ten up to 10^{12}; much larger numbers are given in the epic *Rāmāyaṇa* and the Jaina and Buddhist literature. For a few examples, see R.C. Gupta, *World's longest lists of decuple terms*, Gaṇita Bhāratī Vol **23** (2001), p 83–90.

[60] Legendre was referring to ancient Greeks; he was not aware of ancient Indian research in number theory.

[61] Ancient Indians had the notion of *upapatti* which corresponds to "proof"; see the article by M.D. Srinivas ([Srn]) for a comparison of the two concepts.

> most serious handicaps was the lack of the subscript notation; this was introduced later by Leibniz, in still imperfect form, and its use did not become general until rather late in the next century. Faced with a difficult or complicated proof, Fermat might well be content with a careful analysis of some typical numerical cases, convincing himself at the same time that the steps involved were of general validity. Actually he never succeeded in writing down his proofs.

Now the handicaps of Fermat applied all the more to Jayadeva, Bhāskara II and their predecessors who had no better writing-aid than perishable palm-leaves. Also note that the theoretical justification for *cakravāla* is more complicated. Besides, as mentioned at the beginning of section 5, the great *Ācārya*s (Masters) in ancient Indian tradition usually stated only the essence of their results (and that too in verse form). Details were left for oral transmission; but the unfortunate break in oral tradition, brought about by foreign invasions, resulted in the loss of significant links in mathematical and scientific knowledge. But the very nature of the *cakravāla* (and the term itself) suggests that, like Euler, the Indians must have had a deep insight into the problem if not a complete proof.

The history of ancient Indian algebra, especially that of the *kuṭṭaka*, the *bhāvanā* and the *cakravāla*, is a grand illustration of the following remark of Sri Aurobindo ([A, p 185]):

> Especially in mathematics, astronomy and chemistry, the chief elements of ancient science, she [India] discovered and formulated much and well and anticipated by force of reasoning or experiment some of the scientific ideas and discoveries which Europe first arrived at much later, but was able to base more firmly by her new and completer method.

8. Brahmagupta as a representative of The Classical Age

The *kuṭṭaka*, the *bhāvanā* and the *cakravāla* are fruits of the Classical Age[62] in Indian history, a period of intellectual flourish, a period of a magnificent efflorescence in myriad fields of human endeavour. To a thinker who is aware of the spirit of the times, the Indian achievements on algebra in general, and quadratic indeterminate equations in particular, do not come as a surprise.[63] Thus a familiarity with the broad features of the cultural history of ancient India is desirable for fresh thinking on any topic in ancient Indian science. A deep insight into the history can be had from Sri Aurobindo's essays in [A].

---

[62]The "Classical Age" usually refers to the Gupta period (beginning from 320 CE), the period often identified as the "Golden Age" of post-Vedic Indian history. Many features of the Classical Age can be seen in the history of the preceding centuries, from the time of Mauryas (from around 300 BCE) at least.

[63]Unfortunately, there have been scholars from both India and the West who have nurtured the image of the ancient Indian mind as being exclusively other-worldly.

The early epoch of the *Veda* and the *Upaniṣads*, luminous with the discovery of the Spirit, was the "heroic creative seed-time" ([A, p 110]) — the first formative period of the deeper and finer aspects of the Indian civilisation. This inspired period of great spiritual outflowering passed away into an age of strong intellectuality. In the words of Sri Aurobindo ([A, p 148]):

> The second or post-Vedic age of Indian civilisation was distinguished by the rise of the great philosophies, by a copious, vivid, many-thoughted, many-sided epic literature, by the beginnings of art and science, by the evolution of vigorous and complex society, by the formation of large kingdoms and empires, by manifold formative activities of all kinds and great systems of living and thinking. Here as elsewhere, in Greece, Rome, Persia, China, this was the age of a high outburst of the intelligence working upon life and the things of the mind . . .
>
> It was a birth time and youth of the seeking intellect. . . .

In the Classical Age, each department of knowledge or activity, once taken up, was pursued to its extreme with an amazing intensity, a powerful, penetrating and scrupulous intelligence. Sri Aurobindo explains ([A, p 402–403]):

> For the third power of the ancient Indian spirit was a strong intellectuality, at once austere and rich, robust and minute, powerful and delicate, massive in principle and curious in detail. . . .
>
> The mere mass of the intellectual production during the period from Asoka well into the Mahomedan epoch is something truly prodigious, as can be seen at once if one studies the account which recent scholarship gives of it, and we must remember that that scholarship as yet only deals with a fraction of what is still lying extant and what is extant is only a small percentage of what was once written and known. . . .
>
> In each subject from the largest and most momentous to the smallest and most trivial there was expended the same all-embracing, opulent, minute and thorough intellectuality . . .

Brahmagupta is a striking embodiment of this vibrant intellectuality. This versatile genius systematises arithmetic, lays the foundation of algebra, takes the amazing leap into Pell's equation, records an ingenious construction of rational cyclic quadrilaterals and the formulae for its area and diagonals, gives second order interpolation formula, not to mention the numerous contributions in astronomy. His approach and attitude to mathematical discoveries reveal the mind and spirit of a great pure mathematician enthusiastically engaged in vigorous pursuit of new results *inspired by a feel for their intrinsic mathematical worth* irrespective of the requirements of immediate applications. His mathematical foresight can be seen in his realisation of the importance of "algebra" *as a distinct branch of mathematics*.

Just as Kālidāsa is the great representative poet of the Classical Age, Brahmagupta is its great representative mathematician; the *bhāvanā* shining as a crest-jewel among his mathematical gems.

9. Varga-prakṛti: Motivation and Allied Issues

Brahmagupta's possible motivation for investigating the difficult equation $Dx^2 + 1 = y^2$ in early seventh century was discussed in chapter 7 of our earlier paper ([Du1, p 101–104]). In this section we shall examine a few passages from the book [Vn] of van der Waerden[64], which represent an approach to the Indian history on Pell's equation that is often made by scholars. van der Waerden asks ([Vn, p 148]):

> Brahmagupta was mainly an astronomer. . . . But why should he be interested in the solution of Pell's equation?

The answer can be found in Brahmagupta's text itself. As mentioned earlier (section 2), towards the end of chapter 18, Brahmagupta gives *sukhamātram* as the *raison d'être* for the algebra problems discussed in the chapter:

> These questions are stated simply for delight.

We mention, in this context, that in a lecture on as broad a topic as *Mathematics as a basic science*, Michael Atiyah[65] observed[66]:

> Number Theory for its own sake, as a great intellectual challenge, has a long history, particularly here in India. Already in the 7th century, Brahmagupta made important contributions to what is now known (incorrectly) as Pell's Equation.

As hinted in section 8, Brahmagupta belonged to a phase of Indian history where one sees an insatiable curiosity and an overflowing joy of creation supported by a vigorous intellectual originality. As Sri Aurobindo observes ([A, p 401]):

> When we look at the past of India, what strikes us next is her stupendous vitality, her inexhaustible power of life and joy of life, her almost unimaginably prolific creativeness.

---

[64]B.L. van der Waerden (1903–96) was a leading Dutch mathematician of the 20th century. The world of mathematics is indebted to him for his influential two-volume treatise *Moderne Algebra* (1930) which became the model for the groups-rings-fields approach of text books on abstract algebra. This book, based on the lectures of Emmy Noether and Emil Artin, made a systematic presentation of a large body of research by Richard Dedekind, David Hilbert, Emmy Noether and Emil Artin. van der Waerden is recognised as having contributed most substantially to the promulgation of the ideas of Emmy Noether, the founder of modern "Abstract Algebra". He also did important research in algebraic geometry and various other branches of mathematics. He was also a prominent researcher in history of astronomy and mathematics. Apart from his long-time interest in Greek mathematics, he had developed a deep interest in Babylonian astronomy and mathematics; his book *Science Awakening* (1954) contains his thoughts on the possible flow and modification of ideas from Babylonians to Pythagoreans.

[65]One of the greatest mathematicians of our time, Michael Atiyah (b. 1929) was awarded the Fields Medal in 1966 and the Abel Prize in 2004, the two highest awards in mathematics.

[66]See Current Science, Dec.1993, p 913.

Brahmagupta's delight in performing new feats in algebra was in tune with the general creative impulse in the cultural atmosphere. After bringing clarity and refinement in Āryabhaṭa's *kuṭṭaka*, the pure mathematician's impulse drove him to take up the harder problem of the Pellian equation. Note that his penchant for number-theoretic problems also finds expression in his chapter on arithmetic where he blends his geometric and algebraic skill to give an ingenious construction of a cyclic quadrilateral whose sides, diagonals, circum-radius and area are all rational and whose diagonals are perpendicular to each other (and other related problems). While stating concrete examples, especially those involving the *varga-prakṛti*, Brahmagupta often used the phrase *kurvannāvatsarād gaṇakaḥ — One who can solve it within a year (is truly a) mathematician*. Clearly he revelled in the challenge posed by the *varga-prakṛti*, the determination of a solution of which was described by Jayadeva ([Shu1, p 14–15]) as being *as difficult as setting a fly against the wind*.

van der Waerden overlooks all such internal evidence. He writes ([Vn, p 148]):

> The Greeks had a motive for occupying themselves with Pell's equation. The Pythagoreans wanted to find rational approximations for what we call $\sqrt{2}$, and they found the "Side- and Diagonal-Numbers". Archimedes needed approximations for $\sqrt{3}$, and he used for this purpose solutions of the equations $x^2 = 3y^2 + 1$ and $x^2 = 3y^2 - 2$.

Missing from van der Waerden's discussion is the fact that, right from the Vedic times, there have been efforts by Indian mathematicians to obtain convenient, reasonably accurate, rational approximations to irrational surds $\sqrt{D}$. The approximation $1 + \frac{1}{3} + \frac{1}{3\times 4} - \frac{1}{3\times 4\times 34}\left(= \frac{577}{408}\right)$ for $\sqrt{2}$ was known in the Vedic era and recorded by three authors of the *Śulba-sūtra*s — Baudhāyana, Āpastamba and Kātyāyana.[67] As we mentioned in section 1, the denominator and numerator of the fraction $\frac{577}{408}$ satisfy the Pell equation $2x^2 + 1 = y^2$.[68] A few examples of rational approximations occurring in post-Vedic texts were mentioned in ([Du1, p 102–103]). As mentioned in section 7, the algebraist Nārāyaṇa explicitly used large solutions of $Dx^2 + 1 = y^2$ for approximating $\sqrt{D}$. None of these facts find mention in van der Waerden's book published in 1983. van der Waerden continues ([Vn, p 149; 154]):

> We also have seen that the methods for obtaining these solutions are closely connected with the Euclidean algorithm, applied to irrational ratios of line segments. ... But why should Brahmagupta explain these methods, stripped of their original context, in an astronomical

[67] In as early a text as that of Dickson ([Di, p 341]), the *Śulba* fractions giving approximations to $\sqrt{2}$ are mentioned at the beginning of the chapter "Pell equation".

[68] Also recall (section 7) that the pair (408, 577) is related to the minimal solution (2, 3) by Brahmagupta's identity $(2, 3; 1) \odot (2, 3; 1) \odot (2, 3; 1) \odot (2, 3; 1) = (408, 577; 1)$.

> treatise? I can find only one explanation: he followed an earlier tradition ultimately derived from Greek sources.
>
> I suppose that the Greeks were able to solve Pell's equation, not only for $D = 2$ and $D = 3$, but also for higher values of D, by a systematic application of the Euclidean algorithm. I also suppose that their methods of calculation were copied, without proofs, in Hindu treatises like Brahmagupta's Siddhanta.

Now this is the analysis of a scholar as erudite and meticulous as van der Waerden. What goes wrong? This is not an isolated instance. P. Tannery asserted in 1882 that Indians must have got the solution of Pell's equation from Greeks (who must have solved it before), several historians repeated this assertion in some form or the other, and, a hundred years after Tannery, the writings of van der Waerden (1983) convey a similar impression.[69] Perhaps, as hinted in section 8, consciously or subconsciously, a topic like Pell's equation was considered to be too sophisticated for any ancient mathematician hailing from, what was perceived to be, an other-worldly culture. In this connection, we make a few remarks.

(i) Like Archimedes before him, or Newton and Gauss after him, Brahmagupta was a veritable colossus of his time. This perception is yet to sink in, in the circle of history of mathematics. Anyone familiar with his mathematical gifts would not need to postulate a preceding hypothetical text from where he could have "copied" ("without proof"!) Theorem 1. It is a reflection of the opulent intellectuality and the inexhaustible vital creativeness of the Classical Age that Brahmagupta and some of his successors were doing prolific research in astronomy as well as in pure mathematics with equal vigour and enthusiasm.

(ii) The Indian achievements on Pell's equation tend to overshadow the fact that ancient Indian algebraists had produced a large bulk of work involving ingenious solutions of various types of indeterminate equations (see [DS2] for some examples). Most of these equations were clearly investigated for their own sake. While this zest for exploration of higher degree indeterminate equations (after solving the application-oriented linear case) fits into the general spirit of the era, it was also part of a process of infusing a strong algebra-culture in mathematical thinking. We have seen in section 2 that astronomer-mathematicians like Brahmagupta and Bhāskara II, and mathematicians like Nārāyaṇa, who pursued or discussed the equation $Dx^2 + 1 = y^2$, explicitly emphasised the importance of algebra in their respective texts. Their statements indicate their passion for algebra as also the perception that the cultivation of the subject was a means for sharpening and enriching the intellect.

[69] In ([Sr2, p 17–18]), R. Sridharan quotes similar statements from a paper of van der Waerden as an example of the prejudice of some mathematicians.

Just as Fermat made use of Pell's equation to create interest in number theory, Brahmagupta and Bhāskara II too made use of challenging problems in indeterminate equations (among other topics) to promote algebra among students with promising intellects.

(iii) In the context of the last sentence quoted above from van der Waerden, we note that *proof* (or its absence) can be an issue in the context of *cakravāla*, not *bhāvanā*. Brahmagupta's section *Varga-prakṛti* consists of Theorem 1 which is, in essence, an algebraic identity, and its repeated applications. All these results, *once stated*, can be verified by routine algebraic operations; and, as explained at the beginning of section 5, routine details are not to be expected in a terse (but still voluminous) original ancient Indian treatise. Readers can judge the validity of the quoted speculation by seeing the phrase about some Greek *methods* being copied *without proofs* in Brahmagupta's treatise in the light of the actual contents of the treatise.

(iv) We have seen that Indian history on Pell's equation shows a *gradual development*: first there is *kuṭṭaka*; then the composition law, rational solution and partial integer solution to Pell's equation; and then the complete integer solution (involving all preceding ideas). It is generally agreed that Greek mathematics had stagnated by the 5th–6th century CE; that is before Brahmagupta. If the Greeks had solved the problem before Brahmagupta, and the solution had reached India, why does Brahmagupta record (or "copy") only a partial solution? Or should one presume that, by some combination of circumstances, only an initial part of the hypothetical Greek treatment on Pell's equation had reached Brahmagupta, and a subsequent part reached Jayadeva after a few centuries?

We mention here that it is only in certain aspects of astrology and astronomy that historians have traced a possible transmission of ideas from Greek (or Hellenistic) schools to the Indian treatises. There is no influence of the "pure mathematics" of the Greeks in the developments in post-Vedic Indian mathematics. Now if there has not been transmission of something as standard as Greek geometry[70], what were the chances of transmission of a topic as esoteric (for the times) as the integer solutions of the Pellian equation from the hypothetical Greek sources (and that too in at least two phases — one for Brahmagupta, one for Jayadeva)?

On the one hand, there is concrete textual evidence of research on Pell's equation *progressing in phases* in ancient India in an environment with a good system of enumeration and a clarity regarding formation and handling of equations, with

[70]Indians had a strong culture of systematisation of various branches of knowledge. Mathematicians belonging to an intellectual environment that was used to the structure of Pāṇini's *Aṣṭādhyāyī* would have found the postulational style of Greek geometry suited to their temperament, had they got a glimpse of it. Someone like Brahmagupta, who could come up with gems on the cyclic quadrilateral, would have much more geometry in his treatise if only Greek geometry had reached him. The efforts of Mahāvīra show that Indians did have curiosity regarding conics.

a love for large numbers and a love for algebra (cf. (vi) and (vii) of section 7). On the other, there is as yet no concrete evidence of any Greek mathematician even investigating integer solutions of Pell's equation for large D and no serious evidence of transmission to India of mathematical ideas that were not intimately connected to astronomy. And yet, there has been a tendency to deny the originality of Brahmagupta and his successors regarding their work on *varga-prakṛti*, or to label it, euphemistically, as "disputed". As H.C. Williams recently put it ([Wi, p 404]):

> Finally, there is the belief, apparently due to Tannery, that the cyclic method derives from Greek influences. There seems, apart from possible wishful thinking on the part of Tannery, to be little solid evidence in support of this. The simple fact is that, as mentioned earlier, we don't really know what the Greeks knew about the Pell equation. What we do know, however, is that the Indian methods display a history of steady development and refinement up to and including the discovery of the cyclic method, and this very strongly suggests that Hankel's position that the Indians evolved the technique by themselves is the correct one.

It is to be hoped that the seminar series on History of Mathematics at the Chennai Mathematical Institute will be a precursor to more such congregations in India of leading mathematicians and historians of mathematics, and that such interactions will pave the way for a more accurate, more comprehensive, more representative and more insightful account on history of mathematics in the near future.

Appendix: The Original Verses

To save space, we refrain from quoting original verses for all the statements from ancient Indian treatises referred to in our article. We quote only the earliest extant verses which describe the three principles *kuṭṭaka*, *bhāvanā* and *cakravāla*. English translations for the quoted verses can be found in [ShuS, p 74–84] and [Ke, p 128] (for *kuṭṭaka*), [Du1, p 107–112] (for *bhāvanā*) and [Shu1, p 10–11] (for *cakravāla*). As mentioned at the beginning of section 5, most of these verses are very terse. Faithful translations in English usually require substantial supplementary explanatory notes, apart from a liberal use of brackets for the translations to make sense. Among the famous ancient Indian treatises, perhaps the most lucid exposition of the three principles occurs in the *Bījagaṇita* of Bhāskara II.

Āryabhaṭa's verses on Kuṭṭaka

adhikāgra bhāgahāraṁ chindyādūnāgra bhāgahāreṇa
śeṣaparaspara bhaktaṁ matiguṇamagrāntare kṣiptam
adha upari guṇitamantyayugūnāgra ccheda bhājite śeṣam
adhikāgra ccheda guṇaṁ dvi cchedāgram adhikāgrayutam

Brahmagupta's verses on Bhāvanā

mūlaṁ dvidheṣṭavargād guṇakaguṇādiṣṭa yutavihīnācca
ādyavadho guṇakaguṇaḥ sahāntyaghātena kṛtamantyam
vajravadhaikyaṁ prathamaṁ prakṣepaḥ kṣepavadhatulyaḥ

Jayadeva's verses on Cakravāla

hrasvajyeṣṭhakṣepān pratirāśya kṣepabhaktayoḥ kṣepāt
kuṭṭākāre ca kṛte kiyadguṇaṁ kṣepakaṁ kṣiptvā
tāvatkṛteḥ prakṛtyā hīne prakṣepakeṇa saṁbhakte
svalpatarāvāptiḥ syādityākalito'paraḥ kṣepaḥ
prakṣiptaprakṣepakakuṭṭākāre kaniṣṭhamūlahate
sajyeṣṭhapade prakṣep(ak)eṇa labdhaṁ kaniṣṭhapadam
kṣiptakṣepakakuṭṭāguṇitāttasmātkaniṣṭhamūlahatam
pāscātyaṁ prakṣepaṁ viśodhya śeṣaṁ mahanmūlam
kuryāt kuṭṭākāraṁ punaranayoḥ kṣepabhaktayoḥ padayoḥ
tatseṣṭahatakṣepe sadṛśagune'smin prakṛtihīne
prakṣepaḥ kṣepāpte prakṣiptakṣepakācca guṇakārāt
alpaghnāt sajyeṣṭhāt kṣepāvāptaṁ kaniṣṭhapadam
etatkṣiptakṣepakakuṭṭakaghātādanantarakṣepam
hitvā'lpahataṁ śeṣaṁ jyeṣṭhaṁ tebhyaśca guṇakādi
kuryāttāvadyāvat saṇṇāmekadvicaturṇām patati
iti cakravālakaraṇe'vasaraprāptāni yojyāni

Acknowledgement

The author thanks Amit Roy, Avinash Sathaye, Somesh C. Bagchi and Vivek Hiremath for carefully going through the drafts and for stimulating discussions.

References

[A] Sri Aurobindo, *The Foundations of Indian Culture*, Sri Aurobindo Birth Centenary Library (SABCL) Vol **14** (1972), Sri Aurobindo Ashram, Pondicherry; now reprinted under the title *The Renaissance in India and Other Essays on Indian Culture*. The essays originally appeared serially in *Arya* during 1918–21.

[Ab] S.K. Abhyankar (ed), *Bījagaṇita of Bhāskarācārya*, Bhaskaracharya Pratishthana (1980).

[Bg] A.K. Bag, *The Method of Integral Solution of Indeterminate Equations of the Type: $by = ax \pm c$ in Ancient and Medieval India*, Indian J. History of Science **12(1)** (1977), p 1–16.

[Br] E.J. Barbeau, *Pell's Equation*, Springer (2003).

[BrC] S. Barnard and J. Child, *Higher Algebra*, Macmillan (1936); Indian reprint (2001).

[Ca] F. Cajori, *History of Mathematics*, AMS Chelsea (2000); originally Macmillan 1893; 1919.

[Co1] H.T. Colebrooke, *Algebra, with Arithmetic and Mensuration, from the Sanscrit of Brahmegupta and Bhascara*, John Murray, London (1817); reprinted Cosmo Pub (2004); Sharada (2005).

[Co2] H.T. Colebrooke, *Essays on History Literature and Religions of Ancient India* Vol II, Cosmo Pub (1977); reprint of *Miscellaneous Essays* (1837).

[D1] B. Datta, *Nārāyaṇa's Method for Finding Approximate Value of a Surd*, Bulletin of the Calcutta Mathematical Society **23(4)** (1931), p 187–194.

[D2] B. Datta, *The Science of the Sulba*, University of Calcutta (1932); reprint (1991).

[DS1] B. Datta and A.N. Singh, *History of Hindu Mathematics* Part I, Asia Publishing House (1962); reprinted by Bharatiya Kala Prakashan (2004); originally Motilal Banarasidass (1935).

[DS2] B. Datta and A.N. Singh, *History of Hindu Mathematics* Part II, Asia Publishing House (1962); reprinted by Bharatiya Kala Prakashan (2004); originally Motilal Banarasidass (1938).

[Di] L.E. Dickson, *History of the Theory of Numbers* Vol II, Chelsea Publishing Co. NY (1952); reprinted AMS Chelsea (1999); originally Carnegie Institute Publ. **256** (1920).

[Du1] A.K. Dutta, *Brahmagupta's Bhāvanā: Some Reflections*, Contributions to the History of Indian Mathematics (ed G.G. Emch, R. Sridharan, M.D. Srinivas), CHOM **3**, Hindustan Book Agency (2005), p 77–114.

[Du2] A.K. Dutta, *Āryabhaṭa and Axial Rotation of Earth*, Resonance **11(3)** (March 2006), p 51–68; **11(4)** (April 2006), p 56–74; and **11(5)** (May 2006), p 58–72. Available at the Resonance website
http://www.ias.ac.in/resonance

[F] K.V. Fritz, *The Discovery of Incommensurability by Hippasus of Metapontum*, Annals of Mathematics **46(2)** (1945), p 242–264.

[G] J.R. Goldman, *The Queen of Mathematics: A Historically Motivated Guide to Number Theory*, A.K. Peters (1997).

[Ha] S. Hallgren, *Polynomial-Time Quantum Algorithms for Pell's Equation and the Principal Ideal Problem*, Jour. ACM **54(1)** (2007), p 1–19.

[He] T.L. Heath (ed), *Euclid, The Thirteen Books of The Elements* 2nd edition, Dover (1956); originally Cambridge University Press (1925).

[JJ] G.A. Jones and J.M. Jones, *Elementary Number Theory*, Springer (1998).

[Ke] A. Keller (ed), *Expounding the Mathematical Seed* Vol 1, A Translation of Bhāskara I on the Mathematical Chapter of the Āryabhaṭīya, Birkhauser (2006).

[Kn] D.E. Knuth, *The Art of Computer Programming* Vol 2, Addision-Wesley (1981).

[Kr1] A.A. Krishnaswamy Ayyangar, *New Light on Bhaskara's Chakravala or Cyclic Method of solving Indeterminate Equations of the Second Degree in two Variables*, J. Indian Math. Soc. **18** (1929–30), p 225–248.

[Kr2] A.A. Krishnaswamy Ayyangar, *A New Continued Fraction*, Current Science **6** (June 1938), p 602–604.

[Kr3] A.A. Krishnaswamy Ayyangar, *Theory of the Nearest Square Continued Fraction*, J. Mysore Univ **1** (1940–41), p 21–32 and 97–117; available at
http://www.ms.uky.edu/~sohum/AAK/pdf%20files/nearest%20square.pdf

[L] H.W. Lenstra, *Solving the Pell Equation*, Notices AMS **49(2)** (Feb 2002), p 182–192.

[MRW] K. Matthews, J. Robertson and J. White, *Midpoint Criteria for Solving Pell's Equation Using the Nearest Square Continued Fraction*, Math. Comp. (to appear),
http://www.numbertheory.org/pdfs/nscf_midpoint.pdf

[OR] J.J. O'Connor and E.F. Robertson, *Pell's equation*,
http://www.gap-system.org/~history/HistTopics/Pell.html

[Oj] Manuel Ojanguren, *The Witt Group and The Problem of Lüroth*, Dottorato di ricerca in Matematica, ETS EDITRICE PISA, University of Pisa (1990).

[P] V.B. Panicker (ed), *Bhaskaracharya's Bijaganitham*, Bharatiya Vidya Bhavan (2006).

[PNS] K.S. Patwardhan, S.A. Naimpally and S.L. Singh (ed), *Līlāvatī of Bhāskarācārya*, Motilal Banarasidass (2001).

[Sar] T.A. Sarasvati Amma, *Geometry in Ancient and Medieval India*, Motilal Banarasidass, Delhi (1979; 1999).

[Sat] A. Sathaye, *Precalculus with Geometry and Trigonometry*,
http://www.msc.uky.edu/sohum/ma110/text/ma110_fa08.pdf

[Sel] C.O. Selenius, *Rationale of the Chakravāla Process of Jayadeva and Bhāskara II*, Historia Mathematica **2** (1975), p 167–184.

[Sha] R.S. Sharma (ed), *Brāhma Sphuṭa Siddhānta of Brahmagupta*, Indian Institute of Astronomical Research (1966).

[Shu1] K.S. Shukla, *Ācārya Jayadeva, The Mathematician*, Gaṇita **5** (1954), p 1–20.

[Shu2] K.S. Shukla (ed), *Mahā-Bhāskarīya*, Bhāskara I and his works Part II, Lucknow University (1960).

[Shu3] K.S. Shukla (ed), *Laghu-Bhāskarīya*, Bhāskara I and his works Part III, Lucknow University (1963).

[Shu4] K.S. Shukla (ed), *Āryabhaṭīya of Āryabhaṭa with the commentary of Bhāskara I and Someśvara*, Indian National Science Academy (1976).

[ShuS] K.S. Shukla and K.V. Sarma (ed), *Āryabhaṭīya of Āryabhaṭa*, Indian National Science Academy (1976).

[Sig] P. Singh, *Varga-Prakṛti — The Cakravāla method of its solution and the regular continued fractions*, Indian J. History of Science **19**(1) (1983), p 1–17.

[Sih] K.N. Sinha, *Śrīpati: An Eleventh-Century Indian Mathematician*, Historia Mathematica **12** (1985), p 25–44.

[Sm] D.E. Smith, *History of Mathematics* Vol II, Dover (1958); originally 1925.

[Sr1] R. Sridharan, *Ancient Indian contributions to Quadratic Algebra*, Science in the West and India (ed B.V. Subbarayappa and N. Mukunda), Himalaya Publishing House, Bombay (1995), p 280–289.

[Sr2] R. Sridharan, *Mathematics in Ancient and Medieval India*, Contributions to the History of Indian Mathematics (ed G.G. Emch, R. Sridharan, M.D. Srinivas), CHOM **3**, Hindustan Book Agency (2005), p 1–29.

[Srn] M.D. Srinivas, *Proofs in Indian Mathematics*, Contributions to the History of Indian Mathematics (ed G.G. Emch, R. Sridharan, M.D. Srinivas), CHOM **3**, Hindustan Book Agency (2005), p 209–248.

[Srr] M.S. Sriram, *Algorithms in Indian Mathematics*, Contributions to the History of Indian Mathematics (ed G.G. Emch, R. Sridharan, M.D. Srinivas), CHOM **3**, Hindustan Book Agency (2005), p 153–182.

[Vn] B.L. van der Waerden, *Geometry and Algebra in Ancient Civilizations*, Springer (1983).

[Vr] V.S. Varadarajan, *Algebra in Ancient and Modern Times*, TRIM **14**, Hindustan Book Agency (1997); AMS (1998).

[We1] A. Weil, *History of Mathematics: Why and How*, Proceedings of the International Congress of Mathematicians, Helsinki 1978 Vol 1 (ed. O. Lehto), Helsinki (1980), p 227–236.

[We2] A. Weil, *Number Theory: An approach through history from Hammurapi to Legendre*, Birkhauser (1984).

[Wi] H.C. Williams, *Solving the Pell Equation*, Number Theory for the Millennium Vol III (ed M.A. Bennett et al), A.K. Peters (2002), p 397–436.

Amartya Kumar Dutta
Stat-Math Unit
Indian Statistical Institute
203, B.T. Road; Kolkata 700108
E-mail: amartya@isical.ac.in

Development of Calculus in India

K. Ramasubramanian and M. D. Srinivas

In this article we shall present an overview of the development of calculus in Indian mathematical tradition. The article is divided naturally into two parts. In the first part we shall discuss the developments during what may be called the classical period, starting with the work of Āryabhaṭa (c. 499 CE) and extending up to the work Nārāyaṇa Paṇḍita (c. 1350). The work of the Kerala School starting with Mādhava of Saṅgamagrāma (c. 1350), which has a more direct bearing on calculus, will be dealt with in the second part. Here we shall discuss some of the contributions of the Kerala School during the period 1350–1500 as outlined in the seminal Malayalam work *Yuktibhāṣā* of Jyeṣṭhadeva (c. 1530).

PART I : THE CLASSICAL PERIOD

Āryabhaṭa to Nārāyaṇa Paṇḍita (c. 500–1350 CE)

1. Introduction

In his pioneering history of calculus written sixty years ago, Carl Boyer was totally dismissive of the Indian contributions to the conceptual development of the subject.[1] Boyer's historical overview was written around the same time when (i) Ramavarma Maru Thampuran and Akhileswarayyar brought out the first edition of the Mathematics part of the seminal text *Gaṇita-yukti-bhāṣā*, and (ii) C.T. Rajagopal and his collaborators, in a series of pioneering studies, drew attention to the significance of the results and techniques outlined in *Yuktibhāṣā* (and the work of the Kerala School of Mathematics in general), which seem to have been forgotten after the initial notice by Charles Whish in early nineteenth century. These and the subsequent studies have led to a somewhat different perception of the Indian contribution to the development of calculus as may be gleaned from the following quotation from a recent work on the history of mathematics:[2]

[1]C. B. Boyer, *The History of the Calculus and its Conceptual Development*, Dover, New York 1949, pp. 61–62.

[2]L. H. Hodgekin, *A History of Mathematics: From Mesopotamia to Modernity*, Oxford 2005, p. 168.

> We have here a prime example of two traditions whose aims were completely different. The Euclidean ideology of proof which was so influential in the Islamic world had no apparent influence in India (as al-Biruni had complained long before), even if there is a possibility that the Greek tables of 'trigonometric functions' had been transmitted and refined. To suppose that some version of 'calculus' underlay the derivation of the series must be a matter of conjecture.
>
> The single exception to this generalization is a long work, much admired in Kerala, which was known as *Yuktibhāṣā*, by Jyeṣṭhadeva; this contains something more like proofs—but again, given the different paradigm, we should be cautious about assuming that they are meant to serve the same functions. Both the authorship and date of this work are hard to establish exactly (the date usually claimed is the sixteenth century), but it does give explanations of how the formulae are arrived at which could be taken as a version of the calculus.

The Malayalam work *Gaṇita-yukti-bhāṣā* (c. 1530) of Jyeṣṭhadeva indeed presents an overview of the work of Kerala School of mathematicians during the period 1350–1500 CE. The Kerala School was founded by Mādhava (c. 1340–1420), who was followed by the illustrious mathematician-astronomers Parameśvara (c. 1380–1460), his son Dāmodara and the latter's student Nīlakaṇṭha Somayājī (c. 1444–1550). While the achievements of the Kerala School are indeed spectacular, it has now been generally recognised that these are in fact very much in continuation with the earlier work of Indian mathematicians, especially of the *Āryabhaṭan* school, during the period 500–1350 CE.

In the first part of this article, we shall consider some of the ideas and methods developed in Indian mathematics, during the period 500–1350, which have a bearing on the later work of the Kerala School. In particular, we shall focus on the following topics: Mathematics of zero and infinity; iterative approximations for irrational numbers; summation (and repeated summations) of powers of natural numbers; use of second-order differences and interpolation in the calculation of *jyā* or Rsines; the emergence of the notion of instantaneous velocity of a planet in astronomy; and the calculation of the surface area and volume of a sphere.

2. Zero and Infinity

2.1. Background

The *śānti-mantra* of *Īśāvāsyopaniṣad* (of *Śukla-yajurveda*), a text of *Brahmavidyā*, refers to the ultimate absolute reality, the Brahman, as *pūrṇa*, the perfect, complete or full. Talking of how the universe emanates from the Brahman, it states:

> पूर्णमदः पूर्णमिदं पूर्णात्पूर्णमुदच्यते।
> पूर्णस्य पूर्णमादाय पूर्णमेवावशिष्यते॥
>
> That (Brahman) is *pūrṇa*; this (the universe) is *pūrṇa*; [this] *pūrṇa* emanates

from [that] *pūrṇa*; even when *pūrṇa* is drawn out of *pūrṇa*, what remains is also *pūrṇa*.

Pāṇini's *Aṣṭādhyāyī* (c. 500 BCE) has the notion of *lopa* which functions as a null-morpheme. *Lopa* appears in seven *sūtras* of Chapters 1, 3, 7, starting with

अदर्शनं लोपः ।(1.1.60).

Śūnya appears as a symbol in Piṅgala's *Chandaḥ-sūtra* (c. 300 BCE). In Chapter VIII, while enunciating an algorithm for evaluating any positive integral power of 2 in terms of an optional number of squaring and multiplication (duplication) operations, *śūnya* is used as a marker:

रूपे शून्यम्। द्विः शून्ये। (8.29-30).

Different schools of Indian philosophy have related notions such as the notion of *abhāva* in Nyāya School, and the *śūnyavāda* of the Bauddhas.

2.2. Mathematics of zero in *Brāhmasphuṭa-siddhānta* (c. 628 CE) of Brahmagupta

The *Brāhmasphuṭa-siddhānta* (c. 628 CE) of Brahmagupta seems to be the first available text that discusses the mathematics of zero. *Śūnya-parikarma* or the six operations with zero are discussed in the chapter XVIII on algebra (*kuṭṭakā-dhyāya*), in the same six verses in which the six operations with positives and negatives (*dhanarṇa-ṣaḍvidha*) are also discussed. Zero divided by zero is stated to be zero. Any other quantity divided by zero is said to be *taccheda* (that with zero-denominator):[3]

धनयोर्धनमृणमृणयोर्धनर्णयोरन्तरं समैक्यं खम्।
ऋणमैक्यं च धनमृणधनशून्ययोः शून्यम्॥
ऊनमधिकाद्विशोध्यं धनं धनादृणमृणादधिकमूनात्।
व्यस्तं तदन्तरं स्यादृणं धनं धनमृणं भवति॥
शून्यविहीनमृणमृणं धनं धनं भवति शून्यमाकाशम्।
शोध्यं यदा धनमृणादृणं धनाद्वा तदा क्षेप्यम्॥
ऋणमृणधनयोर्घातो धनमृणयोर्धनवधो धनं भवति।
शून्यर्णयोः खधनयोः खशून्ययोर्वा वधः शून्यम्॥
धनभक्तं धनमृणहृतमृणं धनं भवति खं खभक्तं खम्।
भक्तमृणेन धनमृणं धनेन हृतमृणमृणं भवति॥

[3] *Brāhmasphuṭasiddhānta* of Brahmagupta, Ed. with his own commentary by Sudhakara Dvivedi, Benaras 1902, verses 18.30–35, pp. 309–310.

खोद्धृतमृणं धनं वा तच्छेदं खमृणधनविभक्तं वा।
ऋणधनयोर्वर्गः स्वं खं खस्य पदं कृतिर्यत् तत्॥

...[The sum of] positive (*dhana*) and negative (*ṛṇa*), if they are equal, is zero (*kham*). The sum of a negative and zero is negative, of a positive and zero is positive and of two zeros, zero (*śūnya*).

...Negative subtracted from zero is positive, and positive from zero is negative. Zero subtracted from negative is negative, from positive is positive, and from zero is zero (*ākāśa*). ...The product of zero and a negative, of zero and a positive, or of two zeroes is zero.

...A zero divided by zero is zero. ...A positive or a negative divided by zero is that with zero-denominator.

2.3. Bhāskarācārya on *Khahara*

Bhāskarācārya II (c. 1150), while discussing the mathematics of zero in *Bījagaṇita*, explains that infinity (*ananta-rāśī*) which results when some number is divided by zero is called *khahara*. He also mentions the characteristic property of infinity that it is unaltered even if 'many' are added to or taken away from it, in terms similar to the invocatory verse of *Īśāvāsyopaniṣad* mentioned above:[4]

खहरो भवेत् खेन भक्तश्च राशिः॥
द्विघ्नं त्रिहृत् खं खहृतं त्रयं च शून्यस्य वर्गं वद मे पदं च॥
...अयमनन्तो ३/० राशिः खहरः इत्युच्यते।

अस्मिन्विकारः खहरे न राशावपि प्रविष्टेष्वपि निःसृतेषु।
बहुष्वपि स्याल्लयसृष्टिकालेऽनन्तेऽच्युते भूतगणेषु यद्वत्॥

A quantity divided by zero will be (called) *khahara* (an entity with zero as divisor). Tell me ...three divided by zero ...This infinite (*ananta* or that without end) quantity $\frac{3}{0}$ is called *khahara*.

In this quantity, *khahara*, there is no alteration even if many are added or taken out, just as there is no alteration in the Infinite (*ananta*), Infallible (*acyuta*) [Brahman] even though many groups of beings enter in or emanate from [It] at times of dissolution and creation.

2.4. Bhāskarācārya on multiplication and division by zero

Bhāskarācārya while discussing the mathematics of zero in *Līlāvatī*, notes that when further operations are contemplated, the quantity being multiplied by zero should not be changed to zero, but kept as is. Further he states that when the quantity which is multiplied by zero is also divided by zero, then it remains unchanged.

---

[4] *Bījagaṇita* of Bhāskarācārya, Ed. by Muralidhara Jha, Benaras 1927, *Vāsanā* on *Khaṣaḍvidham* 3, p. 6.

He follows this up with an example and declares that this kind of calculation has great relevance in astronomy:[5]

योगे खं क्षेपसमं वर्गादौ खं खभाजितो राशिः।
खहरः स्यात् खगुणः खं खगुणश्चिन्त्यश्च शेषविधौ॥
शून्ये गुणके जाते खं हारश्चेत्पुनस्तदा राशिः।
अविकृत एव ज्ञेयस्तथैव खेनोनितश्च युतः॥

खं पञ्चयुग्भवति किं वद खस्य वर्गं
मूलं घनं घनपदं खगुणाश्च पञ्च।
खेनोद्धृता दश च कः खगुणो निजार्ध-
युक्तस्त्रिभिश्चगुणितः खहृतस्त्रिषष्टिः॥

...अज्ञातो राशिः तस्य गुणः ०। सार्धं क्षेपः १/२। गुणः ३। हरः ०। दृश्यं ६३।
ततो वक्ष्यमाणेन विलोमविधिना इष्टकर्मणा वा लब्धो राशिः १४।
अस्य गणितस्य ग्रहगणिते महानुपयोगः।

...A quantity multiplied by zero is zero. But it must be retained as such when further operations [involving zero] are contemplated. When zero is the multiplier of a quantity, if zero also happens to be a divisor, then that quantity remains unaltered ...
...What is the number which when multiplied by zero, being added to half of itself multiplied by 3 and divided by zero, amounts to sixty-three?
...Either following the inverse process or by choosing a desired number for the unknown ('rule of false position'), the quantity is obtained to be 14. This kind of calculation is of great use in mathematical astronomy.

Bhāskara works out his example as follows:

$$0\left[\left(x+\frac{x}{2}\right)\times\frac{3}{0}\right] = 63$$
$$\frac{3x}{2}\times 3 = 63.$$
$$\text{Therefore,}\quad x = 14. \qquad (1)$$

Bhāskara, it seems, had not fully mastered this kind of "calculation with infinitesimals" as is clear from the following example that he presents in *Bījagaṇita* while solving quadratic equations by eliminating the middle term:[6]

---

[5] *Līlāvatī* of Bhāskarācārya, Ed. by H. C. Bannerjee, Calcutta 1927, *Vāsanā* on verses 45–46, pp. 14–15.
[6] *Bījagaṇita*, cited above, *Vāsanā* on *avyaktavargādi-samīkaraṇam* 5, pp. 63–64.

कः स्वार्धसहितो राशिः खगुणो वर्गितो युतः।
स्वपदाभ्यां खभक्तश्च जाताः पञ्चदशोच्यताम्॥

Say what is the number which when added to half of itself, multiplied by zero, squared and the square being augmented by twice its root and divided by zero, becomes fifteen?

Clearly the problem as stated is

$$\frac{\left[0(x+\frac{x}{2})\right]^2 + 2\times\left[0(x+\frac{x}{2})\right]}{0} = 15. \tag{2}$$

Bhāskara in his *Vāsanā* seems to just cancel out the zeros without paying any heed to the different powers of zero involved. He converts the problem into the equation

$$\left[x+\frac{x}{2}\right]^2 + 2\times\left[x+\frac{x}{2}\right] = 15. \tag{3}$$

Solving this, by the method of elimination of the middle term, Bhāskara obtains the solution $x = 2$. The other solution $(-\frac{10}{3})$ is not noted.

3. Irrationals and iterative approximations

3.1. Background

Baudhāyana-śulva-sūtra gives the following approximation for $\sqrt{2}$:[7]

प्रमाणं तृतीयेन वर्धयेत्तच्च चतुर्थेनात्मचतुस्त्रिंशोनेन। सविशेषः।

The measure [of the side] is to be increased by its third and this [third] again by its own fourth less the thirty-fourth part [of the fourth]. That is the approximate diagonal (*saviśeṣa*).

$$\begin{aligned}\sqrt{2} &\approx 1+\frac{1}{3}+\frac{1}{3.4}-\frac{1}{3.4.34}\\ &= \frac{577}{408}\\ &= 1.4142156.\end{aligned} \tag{4}$$

[7] *Baudhāyanaśulvasūtram* (1.61-2), in *The Śulvasūtras*, Ed. by S. N. Sen and A. K. Bag, New Delhi 1983, p. 19.

The above approximation is accurate to 5 decimal places. *Baudhāyana-śulva-sūtra*—in the context of discussing the problem of circling a square—also gives an approximation for π:[8]

> चतुरश्रं मण्डलं चिकीर्षन्नक्ष्णयार्धं मध्यात्प्राचीमभ्यापातयेत्।
> यदतिशिष्यते तस्य सहतृतीयेन मण्डलं परिलिखेत्।
>
> If it is desired to transform a square into a circle, [a cord of length] half the diagonal of the square is stretched from the centre to the east; with one-third [of the part lying outside] added to the remainder [of the half-diagonal] the [required] circle is drawn.

If a is half-the side of the square, then the radius r of the circle is given by

$$r \approx \left(\frac{a}{3}\right)(2+\sqrt{2}). \tag{5}$$

This corresponds to $\pi \approx 3.0883$.

3.2. Algorithm for square-roots in *Āryabhaṭīya*

The *Āryabhaṭīya* of Āryabhaṭa (c. 499 CE) gives a general algorithm for computing the successive digits of the square root of a number. The procedure, given in the following verse, is elucidated by us via an example:[9]

> भागं हरेदवर्गान्नित्यं द्विगुणेन वर्गमूलेन।
> वर्गाद्वर्गे शुद्धे लब्धं स्थानान्तरे मूलम्॥
>
> Always divide the non-square (even) place by twice the square-root [already found]. Having subtracted the square [of the quotient] from the square (odd) place, the quotient gives the [digit in the] next place in the square-root.

```
         7  5
     ------------
      5  6  2  5
      4  9
    ---------
14)      7  2   (5
         7  0
       ------
            2  5
            2  5
          ------
            0  0
```

3.3. Approximating the square-root of a non-square number

The method for obtaining approximate square-root (*āsanna-mūla*) of a non-square number (*amūlada-rāśi*) is stated explicitly in *Triśatikā* of Śrīdhara (c. 750):[10]

[8] *Baudhāyanaśulvasūtram* (1.58), ibid., p. 19.

[9] *Āryabhaṭīya* of Āryabhaṭa, Ed. by K. S. Shukla and K. V. Sarma, New Delhi 1976, *Gaṇitapāda* 4, p. 36.

[10] *Triśatikā* of Śrīdhara, Ed. by Sudhakara Dvivedi, Varanasi 1899, verse 46, p. 34.

राशेरमूलदस्याहतस्य वर्गेण केनचिन्महता।
मूलं शेषेण विना विभजेद्गुणवर्गमूलेन॥

Multiply the non-square number by some large square number, take the square-root [of the product] neglecting the remainder, and divide by the square-root of the multiplier.

Nārāyaṇa Paṇḍita (c. 1356) has noted that the solutions of *varga-prakṛti* (the so called Pell's equation) can be used to compute successive approximations to the square-root of a non-square number:[11]

मूलं ग्राह्यं यस्य च तद्रूपक्षेपजे पदे तत्र।
ज्येष्ठं ह्रस्वपदेन च समुद्धरेत् मूलमासन्नम्॥

[With the number] whose square-root is to be found as the *prakṛti* and unity as the *kṣepa*, [obtain the greater and smaller] roots. The greater root divided by the lesser root is an approximate value of the square-root.

Nārāyaṇa considers the example

$$10x^2 + 1 = y^2, \tag{6}$$

and gives the approximate values:

$$\sqrt{10} \approx \frac{19}{6}, \frac{721}{228}, \frac{27379}{8658}, \tag{7}$$

which are obtained by successive compositions (*bhāvanā*) of the solution $x = 6,\ y = 9$:[12]

$$228 = (2)(6)(19), \qquad 721 = (10)(6)^2 + (19)^2, \qquad \text{and so on.}$$

3.4. Approximate value of π in *Āryabhaṭīya*

Āryabhaṭa (c. 499) gives the following approximate value for π:[13]

चतुरधिकं शतमष्टगुणं द्वाषष्टिस्तथा सहस्राणाम्।
अयुतद्वयविष्कम्भस्यासन्नो वृत्तपरिणाहः॥

---

[11] *Gaṇitakaumudī* of Nārāyaṇa Paṇḍita, Ed. by Padmakara Dvivedi, Part II, Benaras 1942, verse 10.17, p. 244.

[12] *Bhāvanā* or the rule of composition enunciated by Brahmagupta is the transformation $(X, Y) \to (X^2 + DY^2,\ 2XY)$ which transforms a solution $x = X,\ y = Y$ of the equation $x^2 - Dy^2 = 1$, into another solution with larger values for x, y, which correspond to higher convergents in the continued fraction expansion of $\sqrt{D}$ and thus give better approximations to it.

[13] *Āryabhaṭīya*, cited above, *Gaṇitapāda* 10, p. 45.

> One hundred plus four multiplied by eight and added to sixty-two thousand: This is the approximate measure of the circumference of a circle whose diameter is twenty-thousand.

Thus as per the above verse $\pi \approx \frac{62832}{20000} = 3.1416$.

3.5. Successive doubling of the sides of the circumscribing polygon

It appears that Indian mathematicians (at least in the Āryabhaṭan tradition) employed the method of successive doubling of the sides of a circumscribing polygon—starting from the circumscribing square leading to an octagon, etc.—to find successive approximations to the circumference of a circle. This method has been described in the later Kerala texts *Yuktibhāṣā* (c. 1530) of *Jyeṣṭhadeva* and *Kriyākramakarī* commentary (c. 1535) of Śaṅkara Vāriyar on *Līlāvatī*, of Bhāskara II. The latter cites the verses of Mādhava (c. 1340–1420) in this connection and notes at the end that:[14]

> एवं यावदभीष्टं सूक्ष्मतामापादयितुं शक्यम्।
>
> Thus, one can obtain [an approximation to the circumference of the circle] to any desired level of accuracy.

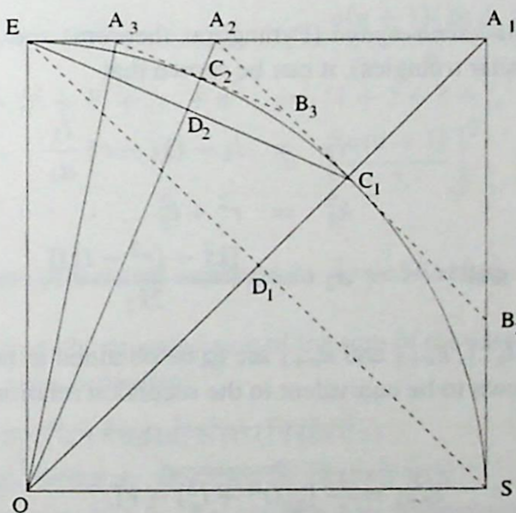


FIGURE 1. Finding the circumference of a square from circumscribing polygons.

[14] *Līlāvatī* of Bhāskara II, Ed. with commentary *Kriyākramakarī* of Śaṅkara Vāriyar by K. V. Sarma, Hoshiarpur 1975, comm. on verse 199, p. 379.

We now outline this method as described in *Yuktibhāṣā*.[15] In Figure 1, $EOSA_1$ is the first quadrant of the square circumscribing the given circle. EA_1 is half the side of the circumscribing square. Let OA_1 meet the circle at C_1. Draw $A_2C_1B_2$ parallel to ES. EA_2 is half the side of the circumscribing octagon.

Similarly, let OA_2 meet the circle at C_2. Draw $A_3C_2B_3$ parallel to EC_1. EA_3 is now half the side of a circumscribing regular polygon of 16 sides. And so on. Let half the sides of the circumscribing square, octagon etc., be denoted

$$l_1 = EA_1,\ l_2 = EA_2,\ l_3 = EA_3,\ \ldots \tag{8}$$

The corresponding *karṇas* (diagonals) are

$$k_1 = OA_1,\ k_2 = OA_2,\ k_3 = OA_3,\ \ldots \tag{9}$$

And the *ābhādhas* (intercepts) are

$$a_1 = D_1A_1,\ a_2 = D_2A_2,\ a_3 = D_3A_3,\ \ldots \tag{10}$$

Now

$$l_1 = r,\ k_1 = \sqrt{2}r \qquad \text{and} \qquad a_1 = \frac{r}{\sqrt{2}}. \tag{11}$$

Using the *bhujā-koṭi-karṇa-nyāya* (Pythagoras theorem) and *trairāśika-nyāya* (rule of three for similar triangles), it can be shown that

$$l_2 = l_1 - (k_1 - r)\,\frac{l_1}{a_1} \tag{12}$$

$$k_2^2 = r^2 + l_2^2 \tag{13}$$

$$\text{and} \qquad a_2 = \frac{[k_2^2 - (r^2 - l_2^2)]}{2k_2}. \tag{14}$$

In the same way l_{n+1}, k_{n+1} and a_{n+1} are to be obtained in terms of l_n, k_n and a_n. These can be shown to be equivalent to the recursion relation:[16]

$$l_{n+1} = \frac{r}{l_n}\,[\sqrt{(r^2 + l_n^2)} - r]. \tag{15}$$

[15] *Gaṇita-yukti-bhāṣā* of Jyeṣṭhadeva, Ed. and Tr. by K. V. Sarma, with Exp. Notes by K. Ramasubramanian, M. D. Srinivas and M. S. Sriram, 2 Vols, Hindustan Book Agency, New Delhi 2008. Reprint Springer 2009, Vol. I Section 6.2, pp. 46–49, 180–83, 366–69.

[16] If we set $r = 1$ and $l_n = \tan\theta_n$, then equation (15) gives $l_{n+1} = \tan\left(\frac{\theta_n}{2}\right)$. Actually, $\theta_n = \frac{\pi}{2^{n+1}}$ and the above method is based on the fact that for large n, $2^n \tan\frac{\pi}{2^{n+2}} \approx 2^n \frac{\pi}{2^{n+2}} = \frac{\pi}{4}$.

4. Summation (and repeated summations) of powers of natural numbers (*saṅkalita*)

4.1. Sum of squares and cubes of natural numbers in *Āryabhaṭīya*

The ancient text *Bṛhaddevatā* (c. 5^{th} century BCE) has the result

$$2 + 3 + \ldots + 1000 = 500,499. \tag{16}$$

Āryabhaṭa (c. 499 CE), in the *Gaṇitapāda* of *Āryabhaṭīya*, deals with a general arithmetic progression in verses 19–20. He gives the sum of the squares and cubes of natural numbers in verse 22:[17]

> सैकसगच्छपदानां क्रमात् त्रिसंवर्गितस्य षष्ठोऽशः।
> वर्गचितिघनः स भवेत् चितिवर्गो घनचितिघनश्च॥
>
> The product of the three quantities, the number of terms plus one, the same increased by the number of terms, and the number of terms, when divided by six, gives the sum of squares of natural numbers (*varga-citi-ghana*). The square of the sum of natural numbers gives the sum of the cubes of natural numbers (*ghana-citi-ghana*).

In other words,

$$1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6} \tag{17}$$

$$\begin{aligned} 1^3 + 2^3 + 3^3 + \ldots + n^3 &= [1 + 2 + 3 + \ldots + n]^2 \\ &= \left[\frac{n(n+1)}{2}\right]^2. \end{aligned} \tag{18}$$

4.2. Repeated sum of natural numbers in *Āryabhaṭīya*

Āryabhaṭa also gives the repeated sum of the sum of the natural numbers (*saṅkalita-saṅkalita* or *vāra-saṅkalita*):[18]

> एकोत्तराद्युपचितेर्गच्छादोकोत्तरत्रिसंवर्गः।
> षड्भक्तः स चितिघनः सैकपदघनो विमूलो वा॥
>
> Of the series (*upaciti*) 1, 2, …, *n*, take three terms in continuation of which the first is the given number of terms (*gaccha*), and find their product; that [product], or the number of terms plus one subtracted from its own cube divided by six, gives the repeated sum (*citi-ghana*).

---

[17] *Āryabhaṭīya*, cited above, *Gaṇitapāda* 22, p. 65.
[18] *Āryabhaṭīya*, cited above, *Gaṇitapāda* 21, p. 64.

We have

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}. \tag{19}$$

Āryabhaṭa's result expresses the sum of these triangular numbers in two forms:

$$\begin{aligned} 1\frac{(1+1)}{2}+2\frac{(2+1)}{2}+\ldots+n\frac{(n+1)}{2} &= \frac{[n(n+1)(n+2)]}{6} \\ &= \frac{[(n+1)^3-(n+1)]}{6}. \end{aligned} \tag{20}$$

4.3. Nārāyaṇa Paṇḍita's general formula for *Vārasaṅkalita*

In his *Gaṇita-kaumudī*, Nārāyaṇa Paṇḍita (c. 1356) gives the formula for the r^{th}-order repeated sum of the sequence of numbers $1, 2, 3, \ldots, n$:[19]

> एकाधिकवारमिताः पदादिरूपोत्तरा पृथक् तेऽंशाः ।
> एकाद्येकचयहरास्तद्घातो वारसङ्कलितम् ॥
>
> The *pada* (number of terms in the sequence) is the first term [of an arithmetic progression] and 1 is the common difference. Take as numerators [the terms in the AP] numbering one more than *vāra* (the number of times the repeated summation is to be made). The denominators are [terms of an AP of the same length] starting with one and with common difference one. The resultant product is *vāra-saṅkalita*.

Let

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}=V_n^{(1)}. \tag{21}$$

Then, Nārāyaṇa's result is

$$V_n^{(r)} = V_1^{(r-1)}+V_2^{(r-1)}+\ldots+V_n^{(r-1)} \tag{22}$$

$$= \frac{[n(n+1)\ldots(n+r)]}{[1.2\ldots(r+1)]}. \tag{23}$$

Nārāyaṇa's result can also be expressed in the form of a sum of polygonal numbers:

$$\sum_{m=1}^{n}\frac{[m(m+1)\ldots(m+r-1)]}{[1.2\ldots r]}=\frac{[n(n+1)\ldots(n+r)]}{[1.2\ldots(r+1)]}. \tag{24}$$

---

[19] *Gaṇitakaumudī* of Nārāyaṇa Paṇḍita, Ed. by Padmakara Dvivedi, Part I, Benaras 1936, verse 3.19–20, p. 123.

This result can be used to evaluate the sums $\sum_{k=1}^{n} k^2,\ \sum_{k=1}^{n} k^3,\ \ldots$ by induction. It can also be used to estimate the behaviour of these sums for large n.

4.4. Summation of geometric series

The geometric series $1 + 2 + 2^2 + \ldots 2^n$ is summed in Chapter VIII of Piṅgala's *Chandaḥ-sūtra* (c. 300 BCE). As we mentioned earlier, Piṅgala also gives an algorithm for evaluating any positive integral power of a number (2 in this context) in terms of an optimal number of squaring and multiplication operations.

Mahāvīrācārya (c. 850), in his *Gaṇita-sāra-saṅgraha* gives the sum of a geometric series and also explains the Piṅgala algorithm for finding the required power of the common ratio between the terms of the series:[20]

पदमितगुणहतिगुणितप्रभवः स्याद्गुणधनं तदाद्यूनम्।
एकोनगुणविभक्तं गुणसङ्कलितं विजानीयात्॥
समदलविषमस्वरूपो गुणगुणितो वर्गताडितो गच्छः।
रूपोनः प्रभवघ्नो व्येकोत्तरभाजितः सारम्॥

The first term when multiplied by the product of the common ratio (*guṇa*) taken as many times as the number of terms (*pada*), gives rise to the *guṇadhana*. This *guṇadhana*, when diminished by the first term and divided by the common ratio less one, is to be understood as the sum of the geometrical series (*guṇa-saṅkalita*).

That is

$$a + ar + ar^2 + \ldots + ar^{n-1} = \frac{a(r^n - 1)}{(r - 1)}. \tag{25}$$

Vīrasena (c. 816), in his commentary *Dhavalā* on the *Ṣaṭkhaṇḍāgama*, has made use of the sum of the following infinite geometric series in his evaluation of the volume of the frustum of a right circular cone:[21]

$$1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \ldots + \left(\frac{1}{4}\right)^n + \ldots = \frac{4}{3}. \tag{26}$$

The proof of the above result is discussed in the *Āryabhaṭīya-bhāṣya* (c. 1502) of Nīlakaṇṭha Somayājī. As we shall see later (section 10.1), Nīlakaṇṭha makes use of this series for deriving an approximate expression for a small arc in terms of the corresponding chord in a circle.

---

[20] *Gaṇitasārasaṅgraha* of Mahāvīrācārya, Ed. by Lakshmi Chanda Jain, Sholapur 1963, verses 2.93–94, pp. 28–29.

[21] See, for instance, T. A. Sarasvati Amma, *Geometry in Ancient and Medieval India*, Motilal Banarsidass, Delhi 1979, Rep. 2007, pp. 203–05.

5. Use of Second-order differences and interpolation in computation of Rsines (*Jyānayana*)

Jyā, *Koṭi* and *Śara*

The *jyā* or *bhujā-jyā* of an arc of a circle is actually the half-chord (*ardha-jyā* or *jyārdha*) of double the arc. In the Figure 2, if R is the radius of the circle, *jyā* (Rsine), *koṭi* or *koṭi-jyā* (Rcosine) and *śara* (Rversine) of the *cāpa* (arc) EC are given by:

$$jy\bar{a}\,(arc\ EC) = CD = R\sin(\angle COE) \quad (27)$$
$$ko\d{t}i\,(arc\ EC) = OD = R\cos(\angle COE) \quad (28)$$
$$\acute{s}ara\,(arc\ EC) = ED = R\text{vers}(\angle COE) = R - R\cos(\angle COE). \quad (29)$$

For computing standard Rsine-tables (*paṭhita-jyā*), the circumference of a circle

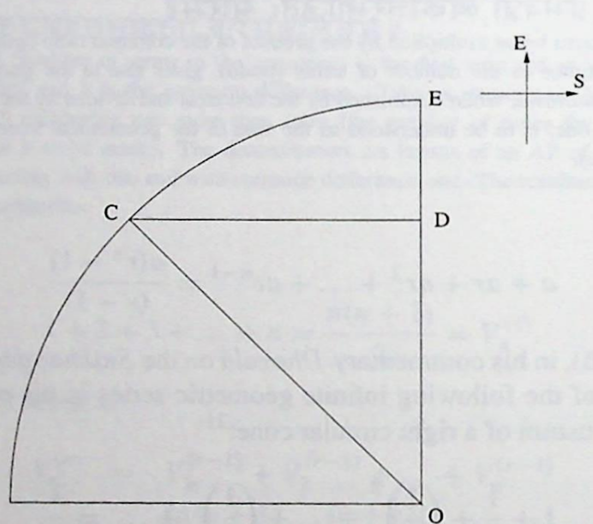


FIGURE 2. *Jyā*, *Koṭi* and *Śara*.

is divided into 21600′ and usually the Rsines are tabulated for every multiple of 225′, thus giving 24 tabulated Rsines in a quadrant. Using the value of $\pi \approx \frac{62832}{20000} = 3.1416$, given by Āryabhaṭa, the value of the radius then turns out to be 3437′ 44″ 19‴. This is accurate up to the seconds, but is usually approximated to 3438′. Using a more accurate value of π, Mādhava (c. 1340–1420) gave the value of the radius correct to the thirds as 3437′ 44″ 48‴ which is also known by the *Kaṭapayādi* formula *devo-viśvasthalī-bhṛguḥ*.

5.1. Computation of Rsines

Once the value of the radius R is fixed (in units of minutes, seconds etc.) the 24 Rsines can be computed (in the same units) using standard relations of *jyotpatti* (trigonometry). For instance, Varāhamihira has given the following Rsine values and relations in his *Pañcasiddhāntikā* (c. 505):[22]

$$R\sin(30^\circ) = \frac{R}{2} \tag{30a}$$

$$R\sin(45^\circ) = \frac{R}{\sqrt{2}} \tag{30b}$$

$$R\sin(60^\circ) = \frac{\sqrt{3}}{2}R \tag{30c}$$

$$R\sin(90^\circ) = R \tag{30d}$$

$$R\sin(A) = R\cos(90-A) \tag{31}$$

$$R\sin^2(A) + R\cos^2(A) = R^2 \tag{32}$$

$$\begin{aligned} R\sin\left(\frac{A}{2}\right) &= \left(\frac{1}{2}\right)[R\sin^2(A) + R\ \text{vers}^2(A)]^{\frac{1}{2}} \\ &= \left(\frac{R}{2}\right)^{\frac{1}{2}}[R - R\cos A]^{\frac{1}{2}}. \end{aligned} \tag{33}$$

The above Rsine values (30) and relations (31)–(33) can be derived using the *bhujā-koṭi-karṇa-nyāya* (Pythagoras theorem) and *trairāśika* (rule of three for similar triangles), as is done for instance in the *Vāsanā-bhāṣya* of Pṛthūdakasvāmin (c. 860) on *Brāhmasphuṭasiddhānta* (c. 628) of Brahmagupta. Equations (30)–(33) can be used to compute all 24 tabular Rsine values.

5.2. Āryabhaṭa's computation of Rsine-differences

The computation of tabular Rsine values was made much simpler by Āryabhaṭa who gave an ingenious method of computing the Rsine-differences, making use of the important property that the second-order differences of Rsines are proportional to the Rsines themselves:[23]

---

[22] *Pañcasiddhāntikā* of Varāhamihira, Ed. by T. S. Kuppanna Sastry and K. V. Sarma, Madras 1993, verses 4.1–5, pp. 76–80.

[23] *Āryabhaṭīya*, cited above, *Gaṇitapāda* 12, p. 51.

प्रथमाद्यापज्यार्धाद्यैरूनं खण्डितं द्वितीयार्धम्।
तत्प्रथमज्यार्धांशैस्तैस्तैरूनानि शेषाणि॥

The first Rsine divided by itself and then diminished by the quotient will give the second Rsine-difference. The same first Rsine, diminished by the quotients obtained by dividing each of the preceding Rsines by the first Rsine, gives the remaining Rsine-differences.

Let $B_1 = R\sin\ (225')$, $B_2 = R\sin\ (450'), \ldots,\ B_{24} = R\sin\ (90^\circ)$, be the twenty-four Rsines, and let $\Delta_1 = B_1,\ \Delta_2 = B_2 - B_1, \ldots,\ \Delta_k = B_k - B_{k-1}, \ldots$ be the Rsine-differences. Then, the above rule may be expressed as[24]

$$\Delta_2 = B_1 - \frac{B_1}{B_1} \tag{34}$$

$$\Delta_{k+1} = B_1 - \frac{(B_1 + B_2 + \ldots + B_k)}{B_1} \qquad (k = 1, 2, \ldots, 23). \tag{35}$$

This second relation is also sometimes expressed in the equivalent form

$$\Delta_{k+1} = \Delta_k - \frac{(\Delta_1 + \Delta_2 + \ldots + \Delta_k)}{B_1} \qquad (k = 1, 2, \ldots, 23). \tag{36}$$

From the above it follows that

$$\Delta_{k+1} - \Delta_k = \frac{-B_k}{B_1} \qquad (k = 1, 2, \ldots, 23). \tag{37}$$

Since Āryabhaṭa also takes $\Delta_1 = B_1 = R\sin(225') \approx 225'$, the above relations reduce to

$$\Delta_1 = 225' \tag{38}$$

$$\Delta_{k+1} - \Delta_k = \frac{-B_k}{225'} \qquad (k = 1, 2, \ldots, 23). \tag{39}$$

5.3. Derivation of the Āryabhaṭa-relation for the second-order Rsine-differences

Āryabhaṭa's relation for the second-order Rsine-differences is derived and made more exact in the *Āryabhaṭīya-bhāṣya* (c. 1502) of Nīlakaṇṭha Somayājī and *Yuktibhāṣā* (c. 1530) of Jyeṣṭhadeva. We shall present a detailed account of

[24]Āryabhaṭa is using the approximation $\Delta_2 - \Delta_1 \approx 1'$ and the second terms in the RHS of (34)–(36) and the RHS of (37) and (39) have an implicit factor of $(\Delta_2 - \Delta_1)$. See (45) below which is exact.

the first and second-order Rsine-differences as given in *Yuktibhāṣā* [25] later in Section 16. Here we shall only summarize the argument.

In Figure 3, the arcs EC_j and EC_{j+1} are successive multiples of 225′. The Rsine and Rcosine of the arcs EC_j and EC_{j+1} are given by

$$B_j = C_j P_j, \; B_{j+1} = C_{j+1} P_{j+1} \tag{40}$$

and

$$K_j = C_j T_j, \; K_{j+1} = C_{j+1} T_{j+1}, \tag{41}$$

respectively. Let M_{j+1} and M_j be the mid-points of the arcs $C_j C_{j+1}$, $C_{j-1} C_j$ and the Rsine and Rcosine of the arcs EM_j and EM_{j+1} be denoted respectively by $B_{j-\frac{1}{2}}$, $B_{j+\frac{1}{2}}$, $K_{j-\frac{1}{2}}$, $K_{j+\frac{1}{2}}$.

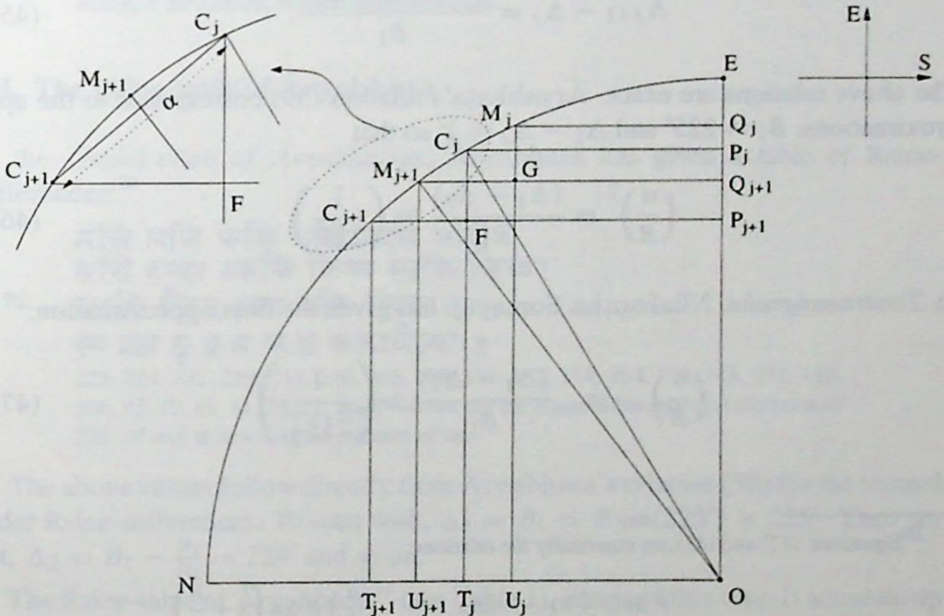


FIGURE 3. Derivation of *Āryabhaṭa* relation.

[25] *Gaṇita-yukti-bhāṣā*, cited above, Section 7.5.1, pp. 94–96, 221–24, 417–20.

Let the chord of the arc C_jC_{j+1}, be denoted by α and let R be the radius. Then a simple argument based on *trairāśika* (similar triangles) leads to the relations:[26]

$$B_{j+1} - B_j = \left(\frac{\alpha}{R}\right) K_{j+\frac{1}{2}} \tag{42}$$

$$K_{j-\frac{1}{2}} - K_{j+\frac{1}{2}} = \left(\frac{\alpha}{R}\right) B_j. \tag{43}$$

Thus we get

$$\begin{aligned} \Delta_{j+1} - \Delta_j &= (B_{j+1} - B_j) - (B_j - B_{j-1}) \\ &= -\left(\frac{\alpha}{R}\right)^2 B_j. \end{aligned} \tag{44}$$

We can also express this relation in the form

$$\Delta_{j+1} - \Delta_j = \frac{-B_j(\Delta_1 - \Delta_2)}{B_1}. \tag{45}$$

The above relations are exact. Āryabhaṭa's relation (39) corresponds to the approximations, $B_1 \approx 225'$ and $\Delta_1 - \Delta_2 \approx 1'$ so that

$$\left(\frac{\alpha}{R}\right)^2 = \frac{(\Delta_1 - \Delta_2)}{B_1} \approx \left(\frac{1}{225'}\right). \tag{46}$$

In *Tantrasaṅgraha*, Nīlakaṇṭha Somayājī has given the finer approximation:[27]

$$\left(\frac{\alpha}{R}\right)^2 = \frac{(\Delta_1 - \Delta_2)}{B_1} \approx \left(\frac{1}{233\frac{1}{2}'}\right). \tag{47}$$

---

[26] Equations (42) and (43) are essentially the relations:

$$\begin{aligned} R\sin(x+h) - R\sin x &= \left(\frac{\alpha}{R}\right) R\cos\left(x + \frac{h}{2}\right) \\ R\cos\left(x - \frac{h}{2}\right) - R\cos\left(x + \frac{h}{2}\right) &= \left(\frac{\alpha}{R}\right) R\sin x, \end{aligned}$$

with $\alpha = 2R\sin\frac{h}{2}$. These lead to (44) in the form:

$$(R\sin(x+h) - R\sin x) - (R\sin x - R\sin(x-h)) = -\left(\frac{\alpha}{R}\right)^2 R\sin x.$$

[27] *Tantrasaṅgraha* of Nīlakaṇṭha Somayājī, Ed. with *Laghu-vivṛti* of Śaṅkara Vāriyar by S. K. Pillai, Trivandrum 1958, verse 2.4, p. 17.

This is further refined by Śaṅkara Vāriyar in his commentary *Laghu-vivṛti* in the form:[28]

$$\left(\frac{\alpha}{R}\right)^2 = \frac{(\Delta_1 - \Delta_2)}{B_1} \approx \left(\frac{1}{233'32''}\right). \tag{48}$$

Since $\alpha = 2R \sin 112'30''$, we find that the above relation is correct up to seconds.

Commenting on Āryabhaṭa's method of computing Rsines, Delambre had remarked:[29]

> The method is curious: it indicates a method of calculating the table of sines by means of their second-differences... This differential process has not up to now been employed except by Briggs, who himself did not know that the constant factor was the square of the chord ΔA $(= 3°45')$ or of the interval, and who could not obtain it except by comparing the second differences obtained in a different manner. The Indians also have probably done the same; they obtained the method of differences only from a table calculated previously by a geometric process. Here then is a method which the Indians possessed and which is found neither amongst the Greeks, nor amongst the Arabs.

5.4. The Rsine-table of Āryabhaṭa

In the *Gītikā-pāda* of *Āryabhaṭīya*, Āryabhaṭa has given a table of Rsine-differences:[30]

> मखि भखि फखि धखि णखि ञखि
> ङखि हस्झ स्ककि किष्ग श्घकि किघ्व।
> घ्लकि किग्र हक्य धकि किच
> स्ग श्झ ङ्व क्ल प्त फ छ कलार्धज्याः ॥
>
> 225, 224, 222, 219, 215, 210, 205, 199, 191, 183, 174, 164, 154, 143, 131, 119, 106, 93, 79, 65, 51, 37, 22, and 7—these are the Rsine-differences [at intervals of 225' of arc] in terms of the minutes of arc.

The above values follow directly from Āryabhaṭa's relation (39) for the second order Rsine-differences. To start with, $\Delta_1 = B_1 = R\sin(225') \approx 225'$. Then we get, $\Delta_2 = B_1 - \frac{B_1}{B_1} = 224'$ and so on.

The Rsine-table of Āryabhaṭa[31] (see Table 1), obtained this way, is accurate up to minutes. In this table, we also give the Rsine values given by Govindasvāmin (c. 825) in his commentary on *Mahābhāskarīya* of Bhāskara I, and by Mādhava

---

[28] Ibid., comm. on verse 2.4.

[29] Delambre, *Historie de l' Astronomie Ancienne*, t 1, Paris 1817, pp. 457, 459f, cited from B. B. Datta and A. N. Singh, 'Hindu Trigonometry', Ind. Jour. Hist. Sc. **18**, 39–108, 1983, p. 77.

[30] *Āryabhaṭīya*, cited above, *Gītikāpāda* 12, p. 29.

[31] See, for instance, A. K. Bag, *Mathematics in Ancient and Medieval India*, Varanasi 1979, pp. 247–48.

(c. 1340–1420) as recorded in the *Āryabhaṭīya-bhāṣya* (c. 1502) of Nīlakaṇṭha Somayājī. Though Govindasvāmin gives the Rsine values up to the thirds, his values are accurate only up to the seconds; those of Mādhava are accurate up to the thirds.

Arc-length	Aryabhaṭa (c. 499)	Govindasvāmin (c. 825)	Mādhava (c. 1375)
3°45′	225′	224′ 50″ 23‴	224′ 50″ 22‴
7°30′	449′	448′ 42″ 53‴	448′ 42″ 58‴
11°15′	671′	670′ 40″ 11‴	670′ 40″ 16‴
15°00′	890′	889′ 45″ 08‴	889′ 45″ 15‴
18°45′	1105′	1105′ 01″ 30‴	1105′ 01″ 39‴
22°30′	1315′	1315′ 33″ 56″	1315′ 34″ 07‴
26°15′	1520′	1520′ 28″ 22‴	1520′ 28″ 35‴
30°00′	1719′	1718′ 52″ 10‴	1718′ 52″ 24‴
33°45′	1910′	1909′ 54″ 19‴	1909′ 54″ 35‴
37°30′	2093′	2092′ 45″ 46‴	2092′ 46″ 03‴
41°15′	2267′	2266′ 38″ 44‴	2266′ 39″ 50‴
45°00′	2431′	2430′ 50″ 54‴	2430′ 51″ 15‴
48°45′	2585′	2584′ 37″ 43‴	2584′ 38″ 06‴
52°30′	2728′	2727′ 20″ 29‴	2727′ 20″ 52‴
56°15′	2859′	2858′ 22″ 31‴	2858′ 22″ 55‴
60°00′	2978′	2977′ 10″ 09‴	2977′ 10″ 34‴
63°45′	3084′	3083′ 12″ 51‴	3083′ 13″ 17‴
67°30′	3177′	3176′ 03″ 23‴	3176′ 03″ 50‴
71°15′	3256′	3255′ 17″ 54‴	3255′ 18″ 22‴
75°00′	3321′	3320′ 36″ 02‴	3320′ 36″ 30‴
78°45′	3372′	3371′ 41″ 01‴	3371′ 41″ 29‴
82°30′	3409′	3408′ 19″ 42‴	3408′ 20″ 11‴
86°15′	3431′	3430′ 22″ 42‴	3430′ 23″ 11‴
90°00′	3438′	3437′ 44″ 19‴	3437′ 44″ 48‴

TABLE 1. Rsine-table of Āryabhaṭa, Govindasvāmin and Mādhava.

5.5. Brahmagupta's second-order interpolation formula

The Rsine table of Āryabhaṭa gives only the Rsine values for the twenty-four multiples of 225′. The Rsines for arbitrary arc-lengths have to be found by interpolation only. In his *Khaṇḍakhādyaka* (c. 665), Brahmagupta gives a second-order interpolation formula for the computation of Rsines for arbitrary arcs. In this

work, which is in the form of a manual (*karaṇa*) for astronomical calculations, Brahmagupta uses a simpler Rsine-table which gives Rsines only at intervals of $15°$ or $900'$:[32]

गतभोग्यखण्डकान्तरदलविकलवधात् शतैर्नवभिराप्त्या।
तद्युतिदलं युतोनं भोग्यादूनाधिकं भोग्यम् ॥

Multiply the residual arc after division by 900′ by half the difference of the tabular Rsine difference passed over (*gata-khaṇḍa*) and to be passed over (*bhogya-khaṇḍa*) and divide by 900′. The result is to be added to or subtracted from half the sum of the same tabular sine differences according as this [half-sum] is less than or equal to the Rsine tabular difference to be passed. What results is the true Rsine-difference to be passed over.

Let h be the basic unit of arc in terms of which the Rsine-table is constructed, which happens to be $225'$ in the case of *Āryabhaṭīya*, and $900'$ in the case of *Khaṇḍakhādyaka*. Let the arc for which Rsine is to be found be given by

$$s = jh + \varepsilon \qquad \text{for some } j = 0, 1, \ldots \tag{49}$$

Now $R\sin(jh) = B_j$ are the tabulated Rsines. Then, a simple interpolation (*trairāśika*) would yield

$$\begin{aligned} R\sin(jh+\epsilon) &= B_j + \left(\frac{\varepsilon}{h}\right)(B_{j+1} - B_j) \\ &= R\sin(jh) + \frac{\varepsilon}{h}\Delta_{j+1}. \end{aligned} \tag{50}$$

Instead of the above simple interpolation, Brahmagupta prescribes

$$R\sin(jh+\epsilon) = B_j + \left(\frac{\varepsilon}{h}\right)\left[\left(\frac{1}{2}\right)(\Delta_j + \Delta_{j+1}) \pm \left(\frac{\varepsilon}{h}\right)\frac{(\Delta_j \sim \Delta_{j+1})}{2}\right]. \tag{51}$$

Here, the sign is chosen to be positive if $\Delta_j < \Delta_{j+1}$, and negative if $\Delta_j > \Delta_{j+1}$ (as in the case of Rsine). So Brahmagupta's rule is actually the second-order interpolation formula

$$\begin{aligned} R\sin(jh+\varepsilon) &= R\sin(jh) + \left(\frac{\varepsilon}{h}\right)\left[\left(\frac{1}{2}\right)(\Delta_j + \Delta_{j+1}) - \left(\frac{\varepsilon}{h}\right)\frac{(\Delta_j - \Delta_{j+1})}{2}\right] \\ &= R\sin(jh) + \left(\frac{\varepsilon}{h}\right)\frac{(\Delta_{j+1} + \Delta_j)}{2} + \left(\frac{\varepsilon}{h}\right)^2\frac{(\Delta_{j+1} - \Delta_j)}{2} \\ &= R\sin(jh) + \left(\frac{\varepsilon}{h}\right)\Delta_{j+1} + \left(\frac{\varepsilon}{h}\right)\left[\frac{\varepsilon}{h} - 1\right]\frac{(\Delta_{j+1} - \Delta_j)}{2}. \end{aligned} \tag{52}$$

[32] *Khaṇḍakhādyaka* of Brahmagupta, Ed. by P. C. Sengupta, Calcutta 1941, p. 151.

6. Instantaneous velocity of a planet (*tātkālika-gati*)

6.1. True daily motion of a planet

In Indian Astronomy, the motion of a planet is computed by making use of two corrections: the *manda-saṃskāra* which essentially corresponds to the equation of centre and the *śīghra-saṃskāra* which corresponds to the conversion of the heliocentric longitudes to geocentric longitudes. The *manda* correction for planets is given in terms of an epicycle of variable radius r, which varies in such a way that

$$\frac{r}{K} = \frac{r_0}{R}, \tag{53}$$

where K is the *karṇa* (hypotenuse) or the (variable) distance of the planet from the centre of the concentric and r_0 is the tabulated (or mean) radius of the epicycle in the measure of the concentric circle of radius R.

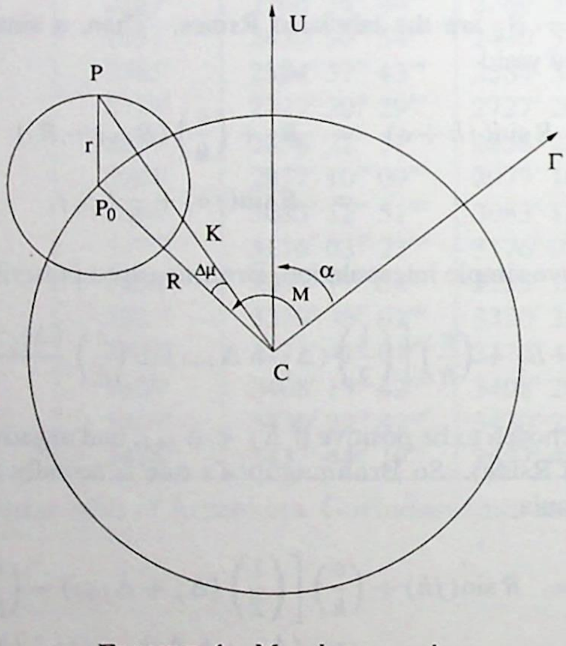


FIGURE 4. *Manda* correction.

In Figure 4, C is the centre of concentric on which the mean planet P_0 is located. CU is the direction of the *ucca* (aphelion or apogee as the case may be). P is the true planet which lies on the epicycle of (variable) radius r centered at P_0, such that P_0P is parallel to CU. If M is the mean longitude of a planet, α the

longitude of the *ucca*, then the correction (*manda-phala*) $\Delta\mu$ is given by

$$\begin{aligned} R\sin(\Delta\mu) &= \left(\frac{r}{K}\right) R\sin(M-\alpha) \\ &= \left(\frac{r_0}{R}\right) R\sin(M-\alpha). \end{aligned} \tag{54}$$

For small r, the left hand side is usually approximated by the arc itself. The *manda*-correction is to be applied to the mean longitude M, to obtain the true or *manda*-corrected longitude μ given by

$$\mu = M - \left(\frac{r_0}{R}\right)\left(\frac{1}{R}\right) R\sin(M-\alpha). \tag{55}$$

If n_m and n_u are the mean daily motions of the planet and the *ucca*, then the true longitude on the next day is given by

$$\mu + n = (M + n_m) - \left(\frac{r_0}{R}\right)\left(\frac{1}{R}\right) R\sin\,(M + n_m - \alpha - n_u). \tag{56}$$

The true daily motion is thus given by

$$n = n_m - \left(\frac{r_0}{R}\right)\left(\frac{1}{R}\right) [R\sin\{(M-\alpha) + (n_m - n_u)\} - R\sin\,(M-\alpha)]. \tag{57}$$

The second term in the above is the correction to mean daily motion (*gati-phala*). An expression for this was given by Bhāskara I (c. 629) in *Mahābhāskarīya*, where he makes use of the approximation:[33]

$$\left.\begin{aligned} R\sin\{(M-\alpha) + (n_m - n_u)\} \\ - R\sin(M-\alpha) \end{aligned}\right\} \approx \left\{\begin{aligned} &(n_m - n_u) \times \\ &\left(\tfrac{1}{225}\right) \text{ Rsine-difference at } (M-\alpha). \end{aligned}\right. \tag{58}$$

In the above approximation, $(n_m - n_u)$ is multiplied by tabular Rsine-difference at the $225'$ arc-bit in which (the tip of the arc) $(M - \alpha)$ is located. Therefore, under this approximation, as long as the anomaly (*kendra*), $(M - \alpha)$, is in the same multiple of $225'$, there will be no change in the *gati-phala* or the correction to the mean velocity. This defect was noticed by Bhāskara also in his later work *Laghubhāskarīya*:[34]

अभिन्नरूपता भुक्तेश्चापभागविचारिणः ।
रवेरिन्दोश्च जीवानामूनभावाद्यसम्भवात् ॥

---

[33] *Mahābhāskarīya* of Bhāskara I, Ed. by K. S. Shukla, Lucknow 1960, verse 4.14, p. 120.

[34] *Laghubhāskarīya* of Bhāskara I, Ed. by K. S. Shukla, Lucknow 1963, verses 2.14-5, p. 6.

एवमालोच्यमानेयं जीवाभुक्तिर्विशीयर्ते।

Whilst the Sun or the Moon moves in the [same] element of arc, there is no change in the rate of motion (*bhukti*), because the Rsine-difference does not increase or decrease; viewed thus, the rate of motion [as given above] is defective.

The correct formula for the true daily motion of a planet, employing the Rcosine as the 'rate of change' of Rsine, seems to have been first given by Muñjāla (c. 932) in his short manual *Laghumānasa* [35] and also by Āryabhaṭa II (c. 950) in his *Mahā-siddhānta*:[36]

कोटिफलघ्नी भुक्तिर्गज्याभक्ता कलादिफलम्॥

The *koṭiphala* multiplied by the [mean] daily motion and divided by the radius gives the minutes of the correction [to the rate of the motion].

This gives the true daily motion in the form

$$n = n_m - (n_m - n_u)\left(\frac{r_0}{R}\right)\left(\frac{1}{R}\right) R\cos(M - \alpha). \qquad (59)$$

6.2. The notion of instantaneous velocity (*tātkālikagati*) according to Bhāskarācārya II

Bhāskarācārya II (c. 1150) in his *Siddhāntaśiromaṇi* clearly distinguishes the true daily motion from the instantaneous rate of motion. And he gives the Rcosine correction to the mean rate of motion as the instantaneous rate of motion. He further emphasizes the fact that the velocity is changing every instant and this is particularly important in the case of Moon because of its rapid motion.[37]

दिनान्तरस्पष्टखगान्तरं स्याद् गतिः स्फुटा तत्समयान्तराले ॥
कोटीफलघ्नी मृदुकेन्द्रभुक्तिस्त्रिज्योद्धृता कर्किमृगादिकेन्द्रे ।
तया युतोना ग्रहमध्यभुक्तिस्तात्कालिकी मन्दपरिस्फुटा स्यात्॥
समीपतिथ्यन्तसमीपचालनं विधोस्तु तत्कालजयैव युज्यते।
सुदूरसञ्चालनमाद्यया यतः प्रतिक्षणं सा न समा महत्यतः ॥

The true daily motion of a planet is the difference between the true planets on successive days. And it is accurate (*sphuṭa*) over that period. The *koṭiphala* (Rcosine of anomaly) is multiplied by the rate of motion of the *manda*-anomaly (*mṛdu-kendra-bhukti*) and divided by the radius. The result added or subtracted from the mean rate of motion of the planet, depending on whether the anomaly is in *Karkyādi* or *Mṛgādi*, gives the true instantaneous rate of motion (*tātkālikī manda-sphuṭagati*) of the planet.

---

[35] *Laghumānasa* of Muñjāla, Ed. by K. S. Shukla, New Delhi 1990, verse 3.4, p. 125.

[36] *Mahāsiddhānta* of Āryabhaṭa II, Ed. by Sudhakara Dvivedi, Varanasi 1910, verse 3.15, p. 58.

[37] *Siddhāntaśiromaṇi* of Bhāskarācārya, Ed. by Muralidhara Chaturvedi, Varanasi 1981, verses 2.36–8, p. 119.

> In the case of the Moon, the ending moment of a *tithi*[38] which is about to end or the beginning time of a *tithi* which is about to begin, are to be computed with the instantaneous rate of motion at the given instant of time. The beginning moment of a *tithi* which is far away can be calculated with the earlier [daily] rate of motion. This is because Moon's rate of motion is large and varies from moment to moment.

Here, Bhāskara explains the distinction between the true daily rate of motion and the true instantaneous rate of motion. The former is the difference between the true longitudes on successive days and it is accurate as the rate of motion, on the average, for the entire period. The true instantaneous rate of motion is to be calculated from the Rcosine of the anomaly (*koṭiphala*) for each relevant moment.

Thus if ω_m and ω_u are the rates of the motion of the mean planet and the *ucca*, then $\omega_m - \omega_u$ is the rate of motion of the anomaly, and the true instantaneous rate of motion of the planet at any instant is given by Bhāskara to be

$$\omega = \omega_m + (\omega_m - \omega_u)\left(\frac{r_0}{R}\right)\left(\frac{1}{R}\right) R\cos(M - \alpha), \qquad (60)$$

where $(M - \alpha)$ is the anomaly of the planet at that instant.

Bhāskara explains the idea of the instantaneous velocity even more clearly in his *Vāsanā*:[39]

> अद्यतनश्वस्तनस्फुटग्रहयोः औदयिकयोर्दिनार्धजयोर्वा अस्तकालिकयोर्वा यदन्तरं कलादिकं सा स्फुटा गतिः। अद्यतनाच्छ्वस्तने न्यूने वक्रागतिर्ज्ञेया। तत्समयान्तराल इति। तस्य कालस्य मध्ये अनया गत्या ग्रहश्चालयितुं युज्यत इति। इयं किल स्थूला गतिः। अथ सूक्ष्मा तात्कालिकी कथ्यते। तुङ्गगत्यूना चन्द्रगतिः केन्द्रगतिः। अन्येषां ग्रहाणां ग्रहगतिरेव केन्द्रगतिः। मृदुकेन्द्रकोटिफलं कृत्वा तेन केन्द्रगतिर्गुण्या। त्रिज्यया भाज्या। लब्धेन कर्क्यादिकेन्द्रे ग्रह-गतिर्युक्ता कार्या। मृगादौ तु रहिता कार्या। एवं तात्कालिकी मन्दपरिस्फुटा स्यात्। तात्कालिक्या भुक्त्या चन्द्रस्य विशिष्टं प्रयो-जनम्। तदाह 'समीपतिथ्यन्तसमीपचालनम्' इति। यत्कालि-कश्चन्द्रः तस्मात् कालाद्गतो वा गम्यो वा यदासन्नस्तितिथ्यन्तस्तदा तात्कालिक्या गत्या तिथिसाधनं कर्तुं युज्यते। तथा समीपचालनं च। यदा तु दूरतरस्तिथ्यन्तो दूरचालनं वा चन्द्रस्य तदाद्यया स्थूलया कर्तुं युज्यते। स्थूलकालत्वात्। यतश्चन्द्रगतिर्महत्वात् प्रतिक्षणं समा न भवति अतस्तदर्थमयं विशेषोऽभिहितः।

---

[38] *Tithi* is the time taken by the Moon to lead the Sun exactly by 12° in longitude.

[39] *Siddhāntaśiromaṇi*, cited above, *Vāsanā* on 2.36–38, p. 119–20.

अथ गतिफलवासना। अद्यतनश्वस्तनग्रहयोरन्तरं गतिः। अत एव ग्रहफलयोरन्तरं गतिफलं भवितुमर्हति। अथ तत्साधनम्। अद्यतनश्वस्तनकेन्द्रयोरन्तरं केन्द्रगतिः। भुजज्याकरणे यद्भोग्यखण्डं तेन सा गुण्या शरद्विदस्रैः (२२५) भाज्या। तत्र तावत् तात्कालिकभोग्यखण्डकरणायानुपातः। यदि त्रिज्यातुल्यया कोटि-ज्ययाद्यं भोग्यखण्डं शरद्विदस्रतुल्यं लभ्यते तदेष्टया किमित्यत्र कोटिज्यायाः शरद्विदस्रा गुणस्त्रिज्या हरः। फलं तात्कालिकं स्फुट-भोग्यखण्डं तेन केन्द्रगतिर्गुणनीया शरद्विदस्रैर्भाज्या।

अत्र शरद्विदस्रमितयोर्गुणकभाजकयोस्तुल्यत्वान्नाशे कृते केन्द्रगतेः कोटिज्यागुणस्त्रिज्याहरः स्यात्। फलमद्यतनश्वस्तनकेन्द्रदोर्ज्ययोरन्तरं भवति। तत्फलकरणार्थं स्वपरिधिना गुण्यं भांशैः (३६०) भाज्यम्। पूर्वं किल गुणकः कोटिज्या सा यावत् परिधिना गुण्यते भांशैः ह्रियते तावत्कोटिफलं जायत इत्युपपन्नं 'कोटीफलघ्नी मृदुकेन्द्रभुक्ति'रित्यादि। एवमद्यतनश्वस्तनग्रहफलयोरन्तरं तद्गतेः फलं कर्क्यादिकेन्द्रे ग्रहर्णफलस्यापचीयमानत्वात् तुलादौ धनफलस्यापचीयमानत्वात् धनम्। मकरादौ तु धनफलस्यापचीयमानत्वात् मेषादावृणफलस्योपचीयमानत्वादृणम् इत्युपपन्नम्।

The true daily velocity is the difference in minutes etc., between the true planets of today and tomorrow, either at the time of sunrise, or mid-day or sunset. If tomorrow's longitude is smaller than that of today, then we should understand the motion to be retrograde. It is said "over that period". This only means that, during that intervening period, the planet is to move with this rate [on the average]. This is only a rough or approximate rate of motion. Now we shall discuss the instantaneous rate of motion... In this way, the *manda*-corrected true instantaneous rate of motion (*tātkālikī manda-parisphuṭagati*) is calculated. In the case of Moon, this instantaneous rate of motion is especially useful...Because of its largeness, the rate of motion of Moon is not the same every instant. Hence, in the case of Moon, the special [instantaneous] rate of motion is stated.

Then, the justification for the correction to the rate of motion (*gati-phala-vāsanā*)...The rate of motion of the anomaly is the difference in the anomalies of today and tomorrow. That should be multiplied by the [current] Rsine-difference used in the computation of Rsines and divided by 225. Now, the following rule of three to obtain the instantaneous Rsine-difference: If the first Rsine-difference 225 results when the Rcosine is equal to the radius, then how much is it for the given Rcosine. In this way, the Rcosine is to be multiplied by 225 and divided by the radius. The result is the instantaneous Rsine-difference and that should be multiplied by the rate of motion in the anomaly and divided by 225...

Thus, Bhāskara is here conceiving also of an instantaneous Rsine-difference, though his derivation of the instantaneous velocity is somewhat obscure. These

ideas are more clearly set forth in the *Āryabhaṭīya-bhāṣya* (c. 1502) of Nīlakaṇṭha Somayājī and other works of the Kerala School.

6.3. The *śīghra* correction to the velocity and the condition for retrograde motion

Bhāskara then goes on to derive the correct expression for the true rate of motion as corrected by the *śīghra*-correction. In the language of modern astronomy, the *śīghra*-correction converts the heliocentric longitude of the planets to the geocentric longitudes. Here also, the Indian astronomers employ an epicycle, but with a fixed radius, unlike in the case of the *manda*-correction.

If μ is the *manda*-corrected (*manda-sphuṭa*) longitude of the planet, ζ is the longitude of the *śīghrocca*, and r_s, the radius of the *śīghra*-epicycle, then the correction (*śīghra-phala*) $\Delta\sigma$ is given by

$$R \sin(\Delta\sigma) = \left(\frac{r_s}{K}\right) \; R \sin\,(\mu - \zeta), \tag{61}$$

where $(\mu - \zeta)$ is the *śīghrakendra* and K is the hypotenuse (*śīghrakarṇa)* given by

$$K^2 = R^2 + r_s^2 - 2Rr_s \cos\,(\mu - \zeta). \tag{62}$$

The calculation of the *śīghra*-correction to the velocity is indeed much more difficult as the denominator in (61), which is the hypotenuse which depends on the anomaly, also varies with time in a complex way. This has been noted by Bhāskara who was able to obtain the correct form of the *śīghra*-correction to the velocity (*śīghra-gati-phala*) in an ingenious way.[40]

> फलांशखाङ्कान्तरशिञ्जिनीघ्नी द्राक्केन्द्रभुक्तिः श्रुतिहृद्विशोध्या।
> स्वशीघ्रभुक्तेः स्फुटखेटभुक्तिः शेषं च वक्रा विपरीतशुद्धौ ॥
>
> The Rsine of ninety degrees, less the degrees of *śīghra*-correction for the longitude (*śīghra-phala*), should be multiplied by the rate of motion of the *śīghra*-anomaly (*drāk-kendra-bhukti*) and divided by the hypotenuse (*śīghra-karṇa*). This, subtracted from the rate of motion of the *śīghrocca*, gives the true velocity of the planet. If this is negative, the planet's motion is retrograde.

If ω is the rate of motion of the *manda*-corrected planet and ω_s is the rate of motion of the *śīghrocca*, then the rate of motion of the *śīghra*-anomaly is $(\omega - \omega_s)$,

---

[40] *Siddhāntaśiromaṇi*, cited above, verse 2.39, p. 121.

and the true velocity of the planet ω_t is given by

$$\omega_t = \omega_s - \left[\frac{(\omega_s - \omega) R \cos(\Delta\sigma)}{K} \right]. \tag{63}$$

The details of the ingenious argument given by Bhāskara for deriving the correct form (63) of the *śīghra*-correction to the velocity has been outlined by D. Arkasomayaji in his translation of *Sīddhāntaśiromaṇi*.[41]

Since Bhāskara's derivation is somewhat long-winded, here we shall present a modern derivation of the result just to demonstrate that the expression given by Bhāskara is indeed exact.

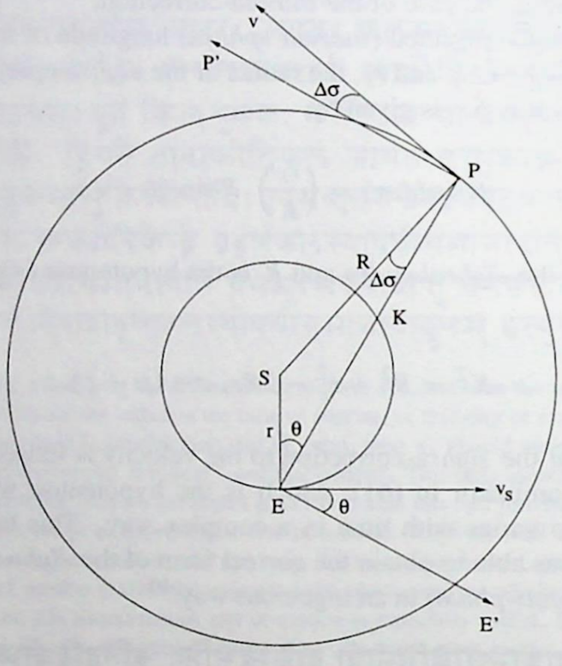


FIGURE 5*a*. Velocity of a planet as seen from the Earth.

In Figure 5*a*, S, E and P represent the positions of the Sun, Earth and an exterior planet respectively. Let v and v_s be the linear velocities of the planet and the Earth with respect to the Sun. PP' and EE' are lines perpendicular to the line EP joining the Earth to the planet. Let R, r represent the radii of the orbits of the planet and the Earth (assumed to be cicular) around the Sun respectively and K, the distance of the planet from the Earth. For an exterior planet, the *śīghra*-correction $\Delta\sigma$ is given by the angle $S\hat{P}E$.

[41]D. Arkasomayaji, *Siddhāntaśiromaṇi* of Bhāskarācārya, Tirupati 1980, pp. 157–161.

If v_t is the linear velocity of the planet as seen from the Earth, then the angular velocity is given by

$$\omega_t = \frac{d\theta}{dt} = \frac{v_t}{K}. \tag{64}$$

The magnitude of v_t in terms of v and v_s (for the situation depicted in Figure 5*a*) is

$$v_t = v \cos \Delta\sigma + v_s \cos\theta. \tag{65}$$

Also, from the triangle SEP, the distance of the planet from the Earth—known as *karṇa*, and denoted K in the figure—may be expressed as

$$\begin{aligned} K &= R\cos\Delta\sigma + r\cos\theta, \\ \text{or} \qquad \cos\theta &= \frac{K - R\cos\Delta\sigma}{r}. \end{aligned} \tag{66}$$

Using (66) in (65) we have

$$\begin{aligned} v_t &= v\cos\Delta\sigma + \frac{v_s}{r}(K - R\cos\Delta\sigma) \\ &= \frac{v_s K}{r} + \cos\Delta\sigma\left(v - v_s\frac{R}{r}\right) \\ \text{or} \qquad \frac{v_t}{K} &= \frac{v_s}{r} + \frac{\cos\Delta\sigma\,\left(v - v_s\frac{R}{r}\right)}{K}. \end{aligned} \tag{67}$$

Making use of (64) and the fact that $v = R\omega$ and $v_s = r\omega_s$, the above equation reduces to

$$\omega_t = \omega_s - \left[\frac{(\omega_s - \omega)R\cos\Delta\sigma}{K}\right],$$

which is same as the expression given by Bhāskara (63).

Bhāskara in his *Vāsanā*[42] explains as to why in the *śīghra* process a different procedure for finding the rate of motion of the planet has to be employed than the one used in the *manda* process:

> अत्रोपपत्तिः। अद्यतनश्वस्तनशीघ्रफलयोरन्तरं गतेः शीघ्रफलं स्यात्।
> तच्च यथा मान्दं गतिफलं ग्रहफलवदानीतं तथा यद्यानीयते कृतेऽपि
> कर्णानुपाते सान्तरमेव स्यात्। यथा धीवृद्धिदे। न हि केन्द्रगतिजमेव

[42]Ibid., *Vāsanā* on 2.39.

फलयोरन्तरं स्यात् किन्त्वन्यदपि अद्यतनभुजफलश्वस्तनभुजफलान्तरे त्रिज्यागुणेऽद्यतनकर्णहृते यादृशं फलं न तादृशं श्वस्तनकर्णहृते। स्वल्पान्तरेऽपि कर्णे भाज्यस्य बहुत्वाद् बह्वन्तरं स्यादित्येतदानयनं हित्वान्यत् महामतिमद्भिः कल्पितम्। तद्यथा...

Here is the justification. The *śīghra*-correction to the rate of motion is the difference between the *śīghra-phalas* of today and tomorrow. If that is derived in the same way as the *manda*-correction to the rate of motion, the result will be incorrect even if it were to be divided by the hypotenuse (*śīghra-karṇa*)... The difference is not just due to the change in the anomaly [which is the argument of the Rsine] but also otherwise... The result of dividing by today's hypotenuse is different from that of dividing by that of tomorrow. Even if the hypotenuses turn out to differ by small amount, the quantities they divide are large and thus a large difference could result. Hence, this way of approach [which was adopted in the case of *manda*-correction to the rate of the motion] has been forsaken and another has been devised by the great intellects. That is as follows...

6.4. The equation of centre is extremum when the velocity correction vanishes

Later, in the *Golādhyāya* of *Siddhāntaśiromaṇi*, Bhāskara considers the situation when the correction to the velocity (*gati-phala*) vanishes:[43]

कक्ष्यामध्यगतिर्यग्रेखाप्रतिवृत्तसंपाते।
मध्यैव गतिः स्पष्टा परं फलं तत्र खेटस्य॥

Where the [North-South] line perpendicular to the [East-West] line of apsides through the centre of the concentric meets the eccentric, there the mean velocity itself is true and the equation of centre is extremum.

In his *Vāsanā*, Bhāskara explains this correlation between vanishing of the velocity correction and the extrema of the correction to the planetary longitude:[44]

कक्ष्यावृत्तमध्ये या तिर्यग्रेखा तस्याः प्रतिवृत्तस्य च यः संपातस्तत्र मध्यैव गतिः स्पष्टा। गतिफलाभावात्। किंच तत्र ग्रहस्य परमं फलं स्यात्। यत्र ग्रहस्य परमं फलं तत्रैव गतिफलाभावेन भवितव्यम्। यतोऽद्यतनश्वस्तनग्रहयोरन्तरं गतिः। फलयोरन्तरं गतिफलम्। ग्रहस्य गतेर्वा फलाभावस्थानमेव धनर्णसन्धिः। यत् पुनर्लल्लोक्तं 'मध्यैव गतिः स्पष्टा वृत्तद्वययोगगे द्युचरे' इति तदसत्। न हि वृत्तद्वययोगे ग्रहस्य परमं फलम्।

The mean rate of motion itself is exact at the points where the line perpendicular [to the line of apsides], at the middle of the concentric circle, meets the eccentric

43 *Siddhāntaśiromaṇi*, cited above, *Golādhyāya* 4.39, p. 393.

44 Ibid., *Vāsanā* on *Golādhyāya* 4.39.

> circle. Because, there is no correction to the rate of motion [at those points]. Also, because there the equation of centre [or correction to the planetary longitudes] is extreme. Wherever the equation of centre is maximum, there the correction to the velocity should be absent. Because, the rate of motion is the difference between the planetary longitudes of today and tomorrow. The correction to the velocity is the difference between the equations of centre. The place where the correction to the velocity vanishes, there is a change over from positive to the negative. And, what has been stated by Lalla, "the mean rate of motion is itself true when the planet is on the intersection of the two circles [concentric and eccentric]", that is incorrect. The planet does not have maximum equation of centre at the confluence of the two circles.

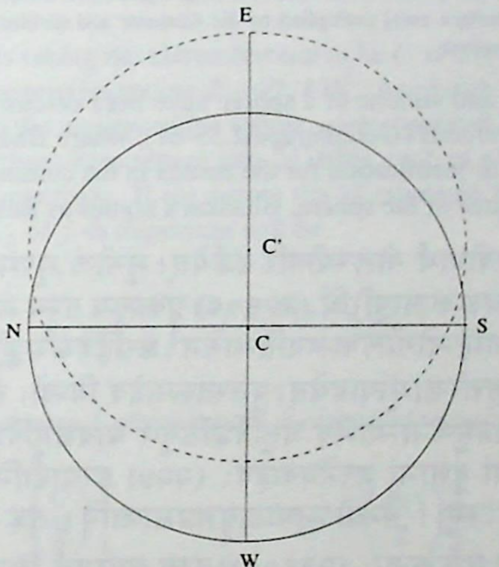


FIGURE 5*b*. Equation of centre is extremum where the correction to velocity vanishes.

Bhāskara explains that when the anomaly is ninety degrees, or the mean planet is at N along the line CN perpendicular to the line of apsides CE (see Figure 5*b*), the equation of centre is maximum. It is precisely then that the correction to the velocity vanishes, as it changes sign from positive to negative. It is incorrect to state (as Lalla did in his *Śiṣyadhīvṛddhida-tantra*) that the correction to the velocity is zero at the point where the concentric and eccentric meet.

7. Surface area and volume of a sphere

In *Āryabhaṭīya* (*Golapāda* 7), the volume of a sphere has been incorrectly estimated as the product of the area of a great circle by its square-root. Śrīdhara

(c. 750) seems to have given the correct expression for the volume of a sphere (*Triśatikā* 56), though his estimate of π is fairly off the mark. Bhāskarācārya (c. 1150) has given the correct relation between the diameter, the surface area and the volume of a sphere in his *Līlāvatī*:[45]

> वृत्तक्षेत्रे परिधिगुणितव्यासपादः फलं यत्
> क्षुण्णं वेदैरुपरि परितः कन्दुकस्येव जालम्।
> गोलस्यैवं तदपि च फलं पृष्ठजं व्यासनिघ्नं
> षड्भिर्भक्तं भवति नियतं गोलगर्भे घनाख्यम्॥
>
> In a circle, the circumference multiplied by one-fourth the diameter is the area, which, multiplied by four, is its surface area going around it like a net around a ball. This [surface area] multiplied by the diameter and divided by six is the volume of the sphere.

The surface area and volume of a sphere have been discussed in greater detail in the *Siddhāntaśiromaṇi* (*Golādhyāya* 2.53-61), where Bhāskara has also presented the *upapatti* or justification for the results in his commentary *Vāsanā*. As regards the surface area of the sphere, Bhāskara argues as follows:[46]

> अथ बालावबोधार्थं गोलस्योपरि दर्शयेत्। भूगोलं मृण्मयं दारुमयं वा कृत्वा तं चक्रकलापरिधिं (२१६००) प्रकल्प्य तस्य मस्तके बिन्दुं कृत्वा तस्माद्बिन्दोर्गोलषण्णवतिभागेन शरद्विदस्रसङ्ख्येन (२२५) धनूरूपेणैव वृत्तरेखामुत्पादयेत्। पुनस्तस्मादेव बिन्दोः तेनैव द्विगुणसूत्रेणान्यां त्रिगुणेनान्यामेवं चतुर्विंशतिगुणं यावच्चतुर्विंशतिर्वृत्तानि भवन्ति। एषां वृत्तानां शरनेत्रबाहवः (२२५) इत्यादीनि ज्यार्धानि व्यासार्धानि स्युः। तेभ्योऽनुपाताद्वृत्तप्रमाणानि। तत्र तावदन्त्यवृत्तस्य मानं चक्रकलाः (२१६००)। तस्य व्यासार्धं त्रिज्या ३४३८। ज्यार्धानि चक्रकलागुणानि त्रिज्याभक्तानि वृत्तमानानि जायन्ते। द्वयोर्द्वयोर्वृत्तयोर्मध्य एकैकं वलयाकारं क्षेत्रम्। तानि चतुर्विंशतिः। बहुज्यापक्षे बहूनि स्युः। तत्र महदधोवृत्तं भूमिमुपरितनं लघुमुखं शरद्विदस्रमितं लम्बं प्रकल्प्य लम्बगुणं कुमुखयोगार्धमित्येवं पृथक् पृथक् फलानि। तेषां फलानां योगो गोलार्धपृष्ठफलम्। तद्द्विगुणं सकलगोलपृष्ठफलम्। तद्व्यासपरिधिघाततुल्यमेव स्यात्।
>
> In order to make the point clear to a beginner, the teacher should demonstrate it on the surface of a sphere. Make a model of the earth in clay or wood and let its circumference be 21, 600 minutes. From the point at the top of the sphere at an arc-distance of $1/96^{th}$ of the circumference, i.e., 225′, draw a circle. Similarly draw circles with twice, thrice,... twenty-four times 225′ [as the arc-distances] so

---

[45] *Līlāvatī*, cited above (fn. 5), verse 203, p. 79–80.

[46] *Siddhāntaśiromaṇi*, cited above, *Vāsanā* on *Golādhyāya* 2.57, p. 362.

> that there will be twenty-four circles. These circles will have as there radii Rsines starting from 225′. The measure [circumference] of the circle will be in proportion to these radii. Here, the last circle has a circumference 21, 600′ and its radius is 3, 438′. The Rsines multiplied by 21, 600 and divided by the radius [3, 438] will give the [circumference] measure of the circles. Between any two circles, there is an annular region and there are twenty-four of them. If more [than 24] Rsines are used, then there will be as many regions. In each figure [if it is cut and spread across as a trapezium] the larger lower circle may be taken as the base and the smaller upper circle as the face and 225′ as the altitude and the area calculated by the usual rule: [Area is] altitude multiplied by half the sum of the base and face. The sum of all these areas is the area of half the sphere. Twice that will be the surface area of the entire sphere. That will always be equal to the product of the diameter and the circumference.

Here Bhāskara is taking the circumference to be $C = 21,600'$, and the corresponding radius is approximated as $R \approx 3,438'$. As shown in Figure 6, circles are drawn parallel to the equator of the sphere, each separated in latitudes by 225′. This divides the northern hemisphere into 24 strips, each of which can be cut and spread across as a trapezium. If we denote the 24 tabulated Rsines by B_1, B_2,... B_{24}, then the area A_j of j-th trapezium will be

$$A_j = \left(\frac{C}{R}\right) \frac{(B_j + B_{j+1})}{2} 225.$$

Therefore, the surface area S of the sphere is estimated to be

$$S = 2\left(\frac{C}{R}\right)\left[B_1 + B_2 + ...B_{23} + \left(\frac{B_{24}}{2}\right)\right](225). \qquad (68)$$

Now, Bhāskara states that the right hand side of the above equation reduces to $2CR$. This can be checked by using Bhāskara's Rsine-table. Bhāskara himself has done the summation of the Rsines in his *Vāsanā* on the succeeding verses,[47] where he gives another method of derivation of the area of the sphere, by cutting the surface of the sphere into lunes. In that context, he computes the sum

$$\begin{aligned} B_1 + B_2.... + B_{23} + \left(\frac{B_{24}}{2}\right) &= B_1 + B_2.... + B_{23} + B_{24} - \left(\frac{R}{2}\right) \\ &\approx 54233 - 1719 = 52514. \end{aligned} \qquad (69)$$

---

[47] *Siddhāntaśiromaṇi*, cited above, *Vāsanā* on *Golādhyāya* 2.58–61, p. 364.

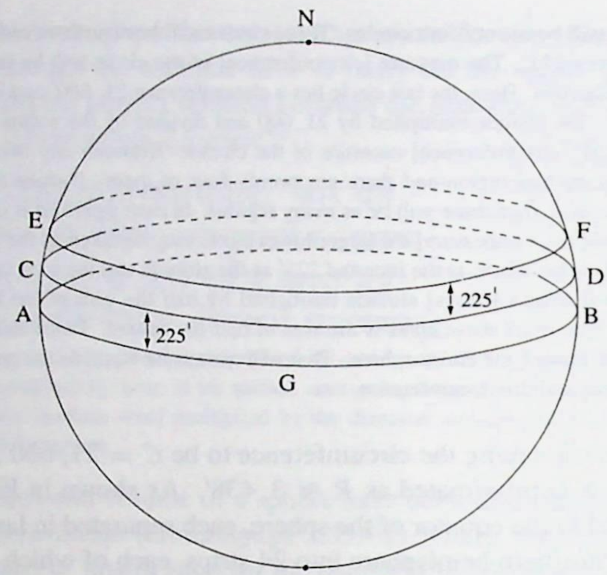


FIGURE 6. Surface area of a sphere.

Thus, according to Bhāskara's Rsine table

$$\left[B_1 + B_2 + ... + B_{23} + \left(\frac{B_{24}}{2}\right)\right](225) = 52514 \times (225)$$
$$= 11815650$$
$$\approx (3437.39)^2. \qquad (70)$$

Taking this as $R^2 = (3438)^2$, we obtain the surface area of the sphere to be [48]

$$S = 2\left(\frac{C}{R}\right)R^2 = 2CR. \qquad (71)$$

Of course, the grossness of the result (70) is due to the fact that the quadrant of the circumference was divided into only 24 bits. Bhāskara also mentions that we may consider dividing the circumference into many more arc-bits, instead of the usual 24 divisions which are made for computing Rsine-tables. This is the approach taken in *Yuktibhāṣā*, where the circumference of the circle is divided into a large

[48] As has been remarked by one of the reviewers, it is indeed intriguing the Bhāskara chose to sum the tabular Rsines numerically, instead of making use of the relation between Rsines and Rcosine-differences which was well known since the time of Āryabhaṭa. In fact, the proof given in *Yuktibhāṣā* (cited below in fn. 49) makes use of the relation between the Rsines and the second order Rsine-differences to estimate this sum.

number, n, of equal arc-bits. If Δ is the Rsine of each arc-bit, the surface area is estimated to be

$$S = 2\left(\frac{C}{R}\right)(B_1 + B_2 + \ldots.B_n)(\Delta). \tag{72}$$

Then it is shown that in the limit of large n,

$$(B_1 + B_2 + \ldots.B_n)(\Delta) \approx R^2, \tag{73}$$

which leads to the result $2CR$ for the surface area.[49]

As regards the volume of a sphere, Bhāskara's justification is much simpler:[50]

> गोलपृष्ठफलस्य व्यासगुणितस्य षडंशो घनफलं स्यात्। अत्रोपपत्तिः। पृष्ठफलसङ्ख्यानि रूपबाहूनि व्यासार्धतुल्यवेधानि सूचीखातानि गोलपृष्ठे प्रकल्प्यानि। सूच्यग्राणां गोलगर्भे संपातः। एवं सूचीफलानां योगो घनफलमित्युपपन्नम्। यत् पुनः क्षेत्रफलमूलेन क्षेत्रफलं गुणितं घनफलं स्यादिति तत् प्रायः चतुर्वेदाचार्यः परमतमुपन्यस्तवान्।
>
> The surface area of a sphere multiplied by its diameter and divided by six is its volume. Here is the justification. As many pyramids as there are units in the surface area with bases of unit side and altitude equal to the semi-diameter should be imagined on the surface of the sphere. The apices of the pyramids meet at the centre of the sphere. Then the volume of the sphere is the sum of the volumes of the pyramids and thus our result is justified. The view that the volume is the product of the area times its own root, is perhaps an alien view (*paramata*) that has been presented by Caturavedācārya [Pṛthūdakasvāmin].

We may note that it is the *Āryabhaṭīya* rule which is referred to as *paramata* in the above passage. Bhāskara's derivation of the volume of a sphere is similar to that of the area of a circle by approximating it as the sum of the areas of a large numbers of triangles with their vertices at the centre, which is actually the proof given in *Yuktibhāṣā*. In the case of the volume of a sphere, *Yuktibhāṣā*, however, gives the more "standard" derivation, where the sphere is divided into a large number of slices and the volume is found as the sum of the volumes of the slices—which ultimately involves estimating the sum of squares of natural numbers (*varga-saṅkalita*), $1^2 + 2^2 + 3^2 + \ldots + n^2$, for large n.[51]

---

[49] *Gaṇita-yukti-bhāṣā*, cited above, Section 7.18, pp. 140–42, 261–63, 465–67. In modern terminology, this amounts to the evaluation of the integral $\int_0^{\frac{\pi}{2}} R\sin\theta R d\theta = R^2$.

[50] *Siddhāntaśiromaṇi*, cited above, *Vāsanā* on *Golādhyāya* 2.61, p. 364.

[51] *Gaṇita-yukti-bhāṣā*, cited above, Section 7.19, pp. 142–45, 263–66, 468–70.

Part II : Work of the Kerala School

Mādhava to Śaṅkara Vāriyar (c. 1350–1550 CE)

8. Kerala School of Astronomy

The Kerala School of Astronomy in the medieval period, pioneered by Mādhava (c. 1340–1420) of Saṅgamagrāma, extended well into the 19th century as exemplified in the work of Śaṅkaravarman (c. 1830), *Rājā* of Kaḍattanaḍu. Only a couple of astronomical works of Mādhava (*Veṇvāroha* and *Sphuṭacandrāpti*) seem to be extant now. Most of his celebrated mathematical discoveries—such as the infinite series for π and the sine and cosine functions—are available only in the form of citations in later works.

Mādhava's disciple Parameśvara (c. 1380–1460) of Vaṭasseri, is reputed to have carried out detailed observations for over 50 years. A large number of original works and commentaries written by him have been published. However, his most important work on mathematics, the commentary *Vivaraṇa* on *Līlāvatī* of Bhāskara II, is yet to be published.

Nīlakaṇṭha Somayājī (c. 1444–1550) of Kuṇḍagrāma, disciple of Parameśvara's son Dāmodara (c. 1410–1520), is the most celebrated member of Kerala School after Mādhava. Nīlakaṇṭha has cited several important results of Mādhava in his various works, the most prominent of them being *Tantrasaṅgraha* (c. 1500) and *Āryabhaṭīya-bhāṣya*. In the latter work, while commenting on *Gaṇitapāda* of *Āryabhaṭīya*, Nīlakaṇṭha has also dealt extensively with many important mathematical issues.

However, the most detailed exposition of the work of the Kerala School, starting from Mādhava, and including the seminal contributions of Parameśvara, Dāmodara and Nīlakaṇṭha, is to be found in the famous Malayalam work *Yuktibhāṣā* (c. 1530) of Jyeṣṭhadeva (c. 1500–1610). Jyeṣṭhadeva was also a disciple of Dāmodara but junior to Nīlakaṇṭha. The direct lineage from Mādhava continued at least till Acyuta Piśāraṭi (c. 1550–1621), a disciple of Jyeṣṭhadeva, who wrote many important works and a couple of commentaries in Malayalam also.

At the very beginning of *Yuktibhāṣā*, Jyeṣṭhadeva states that he intends to present the rationale of the mathematical and astronomical results and procedures which are to be found in *Tantrasaṅgraha* of Nīlakaṇṭha. *Yuktibhāṣā*, comprising 15 chapters, is naturally divided into two parts, Mathematics and Astronomy. Topics in astronomy proper, so to say, are taken up for consideration only from the eighth chapter onwards, starting with a discussion on mean and true planets.

The first seven chapters of *Yuktibhāṣā* are in fact in the nature of an independent treatise on mathematics and deal with various topics which are of relevance to astronomy. It is here that one finds detailed demonstrations of the results of Mādhava such as the infinite series for π, the arc-tangent, sine and the cosine functions, the estimation of correction terms and their use in the generation of faster convergent series. Demonstrations are also provided for the classical results of Āryabhaṭa (c. 499) on *kuṭṭākāra* (linear indeterminate equations), of Brahmagupta (c. 628) on the diagonals and the area of a cyclic quadrilateral, and of Bhāskara II (c. 1150) on the surface area and volume of a sphere. Many of these rationales have also been presented mostly in the form of Sanskrit verses by Śaṅkara Vāriyar (c. 1500–1560) of Tṛkkuṭaveli in his commentaries *Kriyākramakarī* (c. 1535) on *Līlāvatī* of Bhāskara II and *Yukti-dīpikā* on *Tantrasaṅgraha* of Nīlakaṇṭha. In fact, Śaṅkara Vāriyar ends his commentary on the first chapter of *Tantrasaṅgraha* with the acknowledgement:[52]

इत्येषा परक्रोडावासद्विजवरसमीरितो योऽर्थः ।
स तु तन्त्रसङ्ग्रहस्य प्रथमेऽध्याये मया कथितः ॥

Whatever has been the meaning as expounded by the noble *dvija* of *Parakroḍa* [Jyeṣṭhadeva] the same has now been stated by me for the first chapter of *Tantrasaṅgraha*.

In the following sections we shall present an overview of the contribution of the Kerala School to the development of calculus (during the period 1350–1500), following essentially the exposition given in *Yuktibhāṣā*. In order to indicate some of the concepts and methods developed by the Kerala astronomers, we first take up the issue of irrationality of π and the summation of infinite geometric series as discussed by Nīlakaṇṭha Somayājī in his *Āryabhaṭīya-bhāṣya*. We then consider the derivation of binomial series expansion and the estimation of the sum of integral powers of integers, $1^k + 2^k + \ldots + n^k$ for large n, as presented in *Yuktibhāṣā*. These results constitute the basis for the derivation of the infinite series for $\frac{\pi}{4}$ due to Mādhava. We shall outline this as also the very interesting work of Mādhava on the estimation of the end-correction terms and the transformation of the π-series to achieve faster convergence. Finally we shall summarize the derivation of the infinite series for Rsine and Rcosine due to Mādhava.

In the final section, we shall deal with another topic which has a bearing on calculus, but is not dealt with in *Yuktibhāṣā*, namely the evaluation of the instantaneous velocity of a planet. Here, we shall present the result of Dāmodara, as cited by Nīlakaṇṭha, on the instantaneous velocity of a planet which involves

[52] *Tantrasaṅgraha* of Nīlakaṇṭha Somayājī, Ed. with *Yukti-dīpikā* of Śaṅkara Vāriyar by K. V. Sarma, Hoshiarpur 1977, p. 77. The same acknowledgement appears at the end of the subsequent chapters also.

the derivative of the arc-sine function. There are indeed many works and commentaries by later astronomers of the Kerala School, whose mathematical contributions are yet to be studied in detail. We shall here cite only one result due to Acyuta Piṣāraṭi (c. 1550–1621), a disciple of Jyeṣṭhadeva, on the instantaneous velocity of a planet, which involves the evaluation of the derivative of the ratio of two functions.

9. Nīlakaṇṭha's discussion of irrationality of π

In the context of discussing the procedure for finding the approximate square root of a non-square number, by multiplying it by a large square number (the method given in *Triśatikā* of Śrīdhara referred to earlier in Section 3.3), Nīlakaṇṭha observes in his *Āryabhaṭīya-bhāṣya*:[53]

> एवं कृतोऽप्यासन्नमेव मूलं स्यात्। न पुनः करणीमूलस्य तत्त्वतः परिच्छेदः कर्तुं शक्य इत्यभिप्रायः। ततो यावदपेक्षम् अंशानां सूक्ष्मत्वाय महता वर्गेण हननमुक्तम्।
>
> Even if we were to proceed this way, the square root obtained will only be approximate. The idea [that is being conveyed] is, that it is actually not possible to exactly de-limit (*paricchedaḥ*) the square root of a non-square number. Precisely for this reason, multiplication by a large square was stated (recommended) in order to get as much accuracy as desired.

Regarding the choice of the large number that must be made, it is mentioned that one may choose any number—as large a number as possible—that gives the desired accuracy.[54]

> तत्र यावता महता गुणने बुद्धावलंभावः स्यात् तावता हन्यात्। महत्त्वस्य आपेक्षिकत्वात् क्वचिदपि न परिसमाप्तिरिति भावः।
>
> You can multiply by whichever large number you want up to your satisfaction (*buddhāvalaṃbhāvaḥ*). Since largeness is a relative notion, it may be understood that the process is an unending one.

In this context, Nīlakaṇṭha cites the verse given by Āryabhaṭa specifying the ratio of the circumference to the diameter of a circle (value of π), particularly drawing our attention to the fact that Āryabhaṭa refers to this value as "approximate".[55]

---

[53] *Āryabhaṭīya* of Āryabhaṭa, Ed. with *Āryabhaṭīya-bhāṣya* of Nīlakaṇṭha Somayājī by K. Sām-baśiva Śāstrī, Trivandrum Sanskrit Series 101, Trivandrum 1930, comm. on *Gaṇitapāda* 4, p. 14.

[54] Ibid.

[55] Ibid.

वक्ष्यति च – 'अयुतद्वयविष्कम्भस्य आसन्नो वृत्तपरिणाहः' इति। तत्र व्यासेन परिधिज्ञाने अनुमानपरम्परा स्यात्। तत्कर्मण्यपि मूलीकरणस्य अन्तर्भावादेव तस्य आसन्नत्वम्। तत्सर्वं तदवसरे एव प्रतिपादयिष्यामः।

As will be stated [by the author himself] – 'this is [only] an approximate measure of the circumference of the circle whose diameter is twenty-thousand.' In finding the circumference from the diameter, a series of inferences are involved. The approximate nature of this also stems from the fact that it involves finding square roots. All this will be explained later at the appropriate context.

Addressing the issue—later in his commentary, as promised earlier—while discussing the value of π Nīlakaṇṭha observes:[56]

परिधिव्यासयोः सङ्ख्यासम्बन्धः प्रदर्शितः।...आसन्नः, आसन्नतयैव अयुतद्वयसङ्ख्यविष्कम्भस्य इयं परिधिसङ्ख्या उक्ता। कुतः पुनः वास्तवीं सङ्ख्याम् उत्सृज्य आसन्नैव इहोक्ता ? उच्यते। तस्या वक्तुमशक्यत्वात्। कुतः ?

The relation between the circumference and the diameter has been presented. . . . Approximate: This value (62,832) has been stated as only an approximation to the circumference of a circle having a diameter of 20,000. "Why then has an approximate value been mentioned here instead of the actual value?" It is explained [as follows]. Because it (the exact value) cannot be expressed. Why?

Explaining as to why the exact value cannot be presented, Nīlakaṇṭha continues:[57]

येन मानेन मीयमानो व्यासः निरवयवः स्यात्, तेनैव मीयमानः परिधिः पुनः सावयव एव स्यात्। येन च मीयमानः परिधिः निरवयवः तेनैव मीयमानो व्यासोऽपि सावयव एव; इति एकेनैव मानेन मीयमानयोः उभयोः क्वापि न निरवयवत्वं स्यात्। महान्तम् अध्वानं गत्वापि अल्पावयवत्वम् एव लभ्यम्। निरवयवत्वं तु क्वापि न लभ्यम् इति भावः।

Given a certain unit of measurement (*māna*) in terms of which the diameter (*vyāsa*) specified [is just an integer and] has no [fractional] part (*niravayava*), the same measure when employed to specify the circumference (*paridhi*) will certainly have a [fractional] part (*sāvayava*) [and cannot be just an integer]. Again if in terms of certain [other] measure the circumference has no [fractional] part, then employing the same measure the diameter will certainly have a [fractional] part [and cannot be an integer]. Thus when both [the diameter and the circumference] are measured by the same unit, they cannot both be specified [as

[56]Ibid., comm. on *Gaṇitapāda* 10, p. 41.

[57]Ibid., pp. 41–42.

integers] without [fractional] parts. Even if you go a long way (i.e., keep on reducing the measure of the unit employed), the fractional part [in specifying one of them] will only become very small. A situation in which there will be no [fractional] part (i.e, both the diameter and circumference can be specified in terms of integers) is impossible, and this is the import [of the expression *āsanna*].

Evidently, what Nīlakaṇṭha is trying to explain here is the incommensurability of the circumference and the diameter of a circle. Particularly, the last line of the above quote—where Nīlakaṇṭha clearly mentions that, however small you may choose your unit of measurement to be, the two quantities will never become commensurate—is noteworthy.

10. Nīlakaṇṭha's discussion of the sum of an infinite geometric series

In his *Āryabhaṭīya-bhāṣya*, while deriving an interesting approximation for the arc of a circle in terms of the *jyā* (Rsine) and the *śara* (Rversine), Nīlakaṇṭha presents a detailed demonstration of how to sum an infinite geometric series. The specific geometric series that arises in this context is:

$$\frac{1}{4} + \left(\frac{1}{4}\right)^2 + \ldots + \left(\frac{1}{4}\right)^n + \ldots = \frac{1}{3}\,.$$

We shall now present an outline of Nīlakaṇṭha's argument that gives an idea of how the notion of limit was understood in the Indian mathematical tradition.

10.1. Nīlakaṇṭha's approximate formula for the arc in terms of *jyā* and *śara*

In Figure 7, AB is the arc whose length (assumed to be small) is to be determined in terms of the chord lengths AD and BD. In the Indian mathematical literature, the arc AB, the semi-chord AD and the segment BD are referred to as the *cāpa*, *jyārdha* and *śara* respectively. As can be easily seen from the figure, this terminology arises from the fact that these geometrical objects look like a bow, a string and an arrow respectively. Denoting them by c, j, and s, the expression for the arc given by Nīlakaṇṭha may be written as:

$$c \approx \sqrt{\left(1 + \frac{1}{3}\right) s^2 + j^2}. \qquad (74)$$

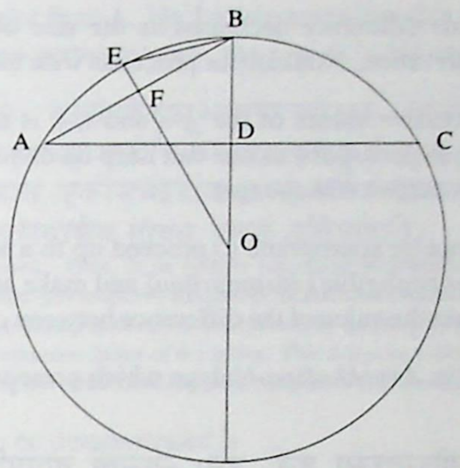


FIGURE 7. Arc-length in terms of *jyā* and *śara*.

Nīlakaṇṭha's proof of the above equation has been discussed in detail by Sarasvati Amma.[58] It may also be mentioned that the above approximation actually does not form a part of the text *Āryabhaṭīya*; but nevertheless it is introduced by Nīlakaṇṭha while commenting upon a verse in *Āryabhaṭīya* that gives the arc in terms of the chords in a circle.[59] The verse that succinctly presents the above equation (74) goes as follows:[60]

सत्र्यंशादिषुवर्गात् ज्यावर्गाढ्यात् पदं धनुः प्रायः ।

The arc is nearly (*prāyaḥ*) equal to the square root of the sum of the square of the *śara* added to one-thirds of it, and the square of the *jyā*.

The proof of (74) given by Nīlakaṇṭha involves:

(1) Repeated halving of the arc-bit, *cāpa* c to get $c_1 \ldots c_i \ldots$.

(2) Finding the corresponding semi-chords, *jyā* (j_i) and the Rversines, *śara* (s_i).

(3) Estimating the difference between the *cāpa* and *jyā* at each step.

If δ_i denotes the difference between the *cāpa* and *jyā* at the i^{th} step, that is,

$$\delta_i = c_i - j_i,$$

---

[58] T. A. Sarasvati Amma, cited above (fn. 21), pp. 179–182.

[59] वृत्ते शरसंवर्गः अर्धज्यावर्गः स खलु धनुषोः ।(*Āryabhaṭīya*, *Gaṇitapāda*, verse 17).

[60] *Āryabhaṭīya-bhāṣya* on *Āryabhaṭīya*, cited above (fn. 50), comm. on *Gaṇitapāda* 12 and 17, p. 63 and p. 110. That the verse cited is from another work of his, namely *Golasāra*, has been alluded to by Nīlakaṇṭha in both the instances of citation.

then it is seen that this difference decreases as the size of the *cāpa* decreases. Having made this observation, Nīlakaṇṭha proceeds with the argument that

- Generating successive values of the j_i-s and s_i-s is an 'unending' process (*na kvacidapi paryavasyati*) as one can keep on dividing the *cāpa* into half *ad infinitum* (*ānantyāt vibhāgasya*).
- It would therefore be appropriate to proceed up to a stage where the difference δ_i becomes negligible (*śūnyaprāya*) and make an 'intelligent approximation', to obtain the value of the difference between c and j approximately.

The original passage in *Āryabhaṭīya-bhāṣya* which presents the above argument reads as follows:[61]

तत्र ज्याचापयोरन्तरस्य पुनः पुनः न्यूनत्वं चापपरिमाणाल्पत्व-क्रमेणेति तत्तदर्धचापानाम् अर्धज्यापरम्परा शरपरम्परा च आनीयमाना न क्वचिदपि पर्यवस्यति आनन्त्याद् विभागस्य। ततः कियन्तञ्चित् प्रदेशं गत्वा चापस्य जीवायाश्च अल्पीयस्त्वम् आपाद्य चापज्यान्तरं च शून्यप्रायं लब्ध्वा पुनरपि कल्प्यमानमन्तरम् अत्यल्पमपि कौशलात् ज्ञेयम्।

10.2. Nīlakaṇṭha's summation of the infinite geometric series

The question that Nīlakaṇṭha poses as he commences his detailed discussion on the sum of geometric series is very important and arises quite naturally whenever one encounters the sum of an infinite series:[62]

कथं पुनः तावदेव वर्धते तावद्वर्धते च ?

How do you know that [the sum of the series] increases only up to that [limiting value] and that it certainly increases up to that [limiting value]?

Proceeding to answer the above question, Nīlakaṇṭha first states the general result

$$a\left[\left(\frac{1}{r}\right)+\left(\frac{1}{r}\right)^2+\left(\frac{1}{r}\right)^3+\ldots\right]=\frac{a}{r-1}.$$

Here, the left hand side is an infinite geometric series with the successive terms being obtained by dividing by a common divisor, r, known as *cheda*, whose value

[61] Ibid., comm. on *Gaṇitapāda* 17, pp. 104–05.
[62] Ibid., p. 106.

is assumed to be greater than 1. He further notes that this result is best demonstrated by considering a particular case, say $r = 4$. In his own words:[63]

> उच्यते। एवं यः तुल्यच्छेदपरभागपरम्परायाः अनन्तायाः अपि संयोगः तस्य अनन्तानामपि कल्प्यमानस्य योगस्य आद्यावयविनः परम्परांशच्छेदात् एकोनच्छेदांशसाम्यं सर्वत्र समानमेव। तद्यथा – चतुरंशपरम्परायामेव तावत् प्रथमं प्रतिपाद्यते।
>
> It is being explained. Thus, in an infinite (*ananta*) geometrical series (*tulya-ccheda-parabhāga-paramparā*) the sum of all the infinite number of terms considered will always be equal to the value obtained by dividing by a factor which is one less than the common factor of the series. That this is so will be demonstrated by first considering the series obtained with one-fourth (*caturaṃśa-paramparā*).

What is intended to be demonstrated is

$$a\left[\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)^2+\left(\frac{1}{4}\right)^3+\ldots\right]=\frac{a}{3}\,. \tag{75}$$

Besides the multiplying factor a, it is noted that, one-fourth and one-third are the only terms appearing in the above equation. Nīlakaṇṭha first defines these numbers in terms of one-twelfth of the multiplier a referred to by the word *rāśi*. For the sake of simplicity we take the *rāśi* to be unity.

$$3\times\frac{1}{12}=\frac{1}{4}; \qquad\qquad 4\times\frac{1}{12}=\frac{1}{3}.$$

Having defined them, Nīlakaṇṭha first obtains the sequence of results,

$$\begin{aligned}
\frac{1}{3} &= \frac{1}{4}+\frac{1}{(4.3)}\,,\\
\frac{1}{(4.3)} &= \frac{1}{(4.4)}+\frac{1}{(4.4.3)}\,,\\
\frac{1}{(4.4.3)} &= \frac{1}{(4.4.4)}+\frac{1}{(4.4.4.3)}\,,
\end{aligned}$$

and so on, which leads to the general result,

$$\frac{1}{3}-\left[\frac{1}{4}+\left(\frac{1}{4}\right)^2+\ldots+\left(\frac{1}{4}\right)^n\right]=\left(\frac{1}{4}\right)^n\left(\frac{1}{3}\right). \tag{76}$$

---

[63] Ibid., pp. 106–07.

Nīlakaṇṭha then goes on to present the following crucial argument to derive the sum of the infinite geometric series: As we sum more terms, the difference between $\frac{1}{3}$ and sum of powers of $\frac{1}{4}$ (as given by the right hand side of (76)), becomes extremely small, but never zero. Only when we take all the terms of the infinite series together do we obtain the equality

$$\frac{1}{4} + \left(\frac{1}{4}\right)^2 + \ldots + \left(\frac{1}{4}\right)^n + \ldots = \frac{1}{3}. \tag{77}$$

A brief extract from the text presenting the above argument is given below:[64]

ये राशेर्द्वादशांशाः तेषां त्रिकं हि चतुरंशः। चतुष्कं च त्र्यंशः। तच्चतुष्टये त्र्यंशात्मके भागत्रयं चतुरंशेनापूर्णम्। यः पुनः तस्य चतुर्थोऽम्शः तस्यापि पादत्रयं चतुरंशस्य चतुरंशेनापूर्णम्। द्वादशांशानां त्रयाणां...

तस्य पुनः पुनरतिसूक्ष्मत्वादेव न केवलं त्र्यंशत्वेन अङ्गीकारः, निरूप्यमाणस्य वा क्रियमाणस्य वा आनन्त्यात्। आनन्त्यादेव शिष्टत्वादेव कर्मणस्तस्य अपरिपूर्तिर्भाति। एवं सर्वदापि सावशेषाणां कर्मणां परम्परायां कार्त्स्न्येनाकृष्यात्र सन्निहितायां परिपूर्तिः स्यादेवेति निश्चीयते चतुर्गुणोत्तरे गुणोत्तराख्ये गणितेऽपि।

Three times one-twelfth of a *rāśi* is one-fourth (*caturaṃśa*) [of that *rāśi*]. Four times that is one-third (*tryaṃśa*). [Considering] four times that [one-twelfth of the *rāśi*] which is one-third, three by fourth of that falls short by one-fourth [of one-third of the *rāśi*]. Three-fourths of that [i.e., of $\frac{1}{4.3}$ of the *rāśi*] which is one-fourth of that (*tryaṃśa*), again falls short [of the same] by one-fourth of one-fourth [of one-third of the *rāśi*] ...

Since the result to be demonstrated or the process to be carried out is never ending (*ānantyāt*) and the difference though very small (*atisūkṣmatvāt*) [still exists and the sum of the series] cannot be simply taken to be one-third. It seems that the process is incomplete since always something remains because of its never ending nature. In fact, since in all the problems involving [infinite] series, by bringing in all the terms and placing them together, the process would [in principle] become complete, here, in the mathematics involving repeated multiplication of one-fourth, a similar conclusion may be drawn.

[64] Ibid., p. 107.

11. Derivation of binomial series expansion

Yuktibhāṣā presents a very interesting derivation of the binomial series for $(1+x)^{-1}$ by making iterative substitutions in an algebraic identity. The method given in the text may be summarized as follows.[65]

Consider the product $a\left(\frac{c}{b}\right)$, where some quantity a is multiplied by the multiplier c, and divided by the divisor b. Here, a is called *guṇya*, c the *guṇaka* and b the *hāra*, which are all assumed to be positive. Now the above product can be rewritten as:

$$a\left(\frac{c}{b}\right) = a - a\,\frac{(b-c)}{b}. \tag{78}$$

In the expression $a\frac{(b-c)}{b}$ in (78) above, if we want to replace the division by b (the divisor) by division by c (the multiplier), then we have to make a subtractive correction (called *śodhya-phala*) which amounts to the following equation.

$$a\frac{(b-c)}{b} = a\frac{(b-c)}{c} - \left(a\frac{(b-c)}{c} \times \frac{(b-c)}{b}\right). \tag{79}$$

Now, in the second term (inside parenthesis) in (79)—which is what we referred to as *śodhya-phala*, which literally means a quantity to be subtracted—if we again replace the division by b by division by c, then we have to employ the relation (79) once again to get another subtractive term

$$\begin{aligned} a\frac{c}{b} &= a - \left[a\frac{(b-c)}{c} - a\frac{(b-c)}{c} \times \frac{(b-c)}{b}\right] \\ &= a - \left[a\frac{(b-c)}{c} - a\frac{(b-c)}{c} \times \frac{(b-c)}{c} \times \frac{c}{b}\right] \\ &= a - \left[a\frac{(b-c)}{c} - \left[a\frac{(b-c)^2}{c^2} - \left(a\frac{(b-c)^2}{c^2} \times \frac{(b-c)}{b}\right)\right]\right]. \end{aligned} \tag{80}$$

Here, the quantity $a\frac{(b-c)^2}{c^2}$ is called *dvitīya-phala* or simply *dvitīya* and the one subtracted from that is *dvitīya-śodhya-phala*. If we carry out the same set of operations, the m^{th} *śodhya-phala* subtracted from the m^{th} term will be of the form

$$a\left[\frac{(b-c)}{c}\right]^m - a\left[\frac{(b-c)}{c}\right]^m \times \frac{(b-c)}{b}.$$

---

[65] *Gaṇita-yukti-bhāṣā*, cited above, Vol. I, Sections 6.3.3–4, pp. 54–58, 188–191, 375–378.

Since the successive *śodhya-phalas* are subtracted from their immediately preceding term, we will end up with a series in which all the odd terms (leaving out the *guṇya*, *a*) are negative and the even ones positive. Thus, after taking *m* *śodhya-phalas* we get

$$a\frac{c}{b} = a - a\frac{(b-c)}{c} + a\left[\frac{(b-c)}{c}\right]^2 - \ldots + (-1)^m a\left[\frac{(b-c)}{c}\right]^m + (-1)^{m+1} a\left[\frac{(b-c)}{c}\right]^m \frac{(b-c)}{b}. \quad (81)$$

Regarding the question of termination of the process, both the texts *Yuktibhāṣā* and *Kriyākramakarī* clearly mention that logically there is no end to the process of generating *śodhya-phalas*. We may thus write our result as:[66]

$$a\frac{c}{b} = a - a\frac{(b-c)}{c} + a\left[\frac{(b-c)}{c}\right]^2 - \ldots + (-1)^{m-1} a\left[\frac{(b-c)}{c}\right]^{m-1} + (-1)^m a\left[\frac{(b-c)}{c}\right]^m + \ldots. \quad (82)$$

It is also noted that the process may be terminated after having obtained the desired accuracy by neglecting the subsequent *phalās* as their magnitudes become smaller and smaller. In fact, *Kriyākramakarī* explicitly mentions the condition under which the succeeding *phalas* will become smaller and smaller:[67]

> एवं मुहुः फलानयने कृतेऽपि युक्तितः क्वापि न समाप्तिः। तथापि यावदपेक्षं सूक्ष्मतामापाद्य पाश्चात्यान्युपेक्ष्य फलानयनं समापनीयम्। इहोत्तरोत्तरफलानां न्यूनत्वं तु गुणहारान्तरे गुणकाराऩ्यून एव स्यात्।
>
> Thus, even if we keep finding the *phalas* repeatedly, logically there is no end to the process. Even then, having carried on the process to the desired accuracy (*yāvadapekṣaṃ sūkṣmatāmāpādya*), one should terminate computing the *phalas* by [simply] neglecting the terms that may be obtained further (*pāścātyānyupekṣya*). Here, the succeeding *phalas* will become smaller and smaller only when the difference between the *guṇaka* and *hāra* is smaller than *guṇaka* [that is $(b \sim c) < c$].

---

[66] It may be noted that if we set $\frac{(b-c)}{c} = x$, then $\frac{c}{b} = \frac{1}{(1+x)}$. Hence, the series (82) is none other than the well known binomial series

$$\frac{a}{1+x} = a - ax + ax^2 - \ldots + (-1)^m ax^m + \ldots,$$

which is convergent for $-1 < x < 1$.

[67] *Kriyākramakarī* on *Līlāvatī*, cited above (fn. 14), comm. on verse 199, p. 385.

12. Estimation of sums of $1^k + 2^k + \ldots n^k$ for large n

As mentioned in section 4.1, Āryabhaṭa has given the explicit formulae for the summation of squares and cubes of integers. The word employed in the Indian mathematical literature for summation is *saṅkalita*. The formulae given by Āryabhaṭa for the *saṅkalitas* are as follows:

$$\begin{aligned} S_n^{(1)} &= 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \\ S_n^{(2)} &= 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \\ S_n^{(3)} &= 1^3 + 2^3 + \cdots + n^3 = \left[\frac{n(n+1)}{2}\right]^2. \end{aligned} \tag{83}$$

From these, it is easy to estimate these sums when n is large. *Yuktibhāṣā* gives a general method of estimating the *sama-ghāta-saṅkalita*

$$S_n^{(k)} = 1^k + 2^k + \cdots + n^k, \tag{84}$$

when n is large. The text presents a general method of estimation, which does not make use of the actual value of the sum. In fact, the same argument is repeated even for $k = 1, 2, 3$, although the result of summation is well known in these cases.

12.1. The sum of natural numbers (*Mūla-saṅkalita*)

Yuktibhāṣā takes up the discussion on *saṅkalitas* in the context of evaluating the circumference of a circle which is conceived to be inscribed in a square. It is half the side of this square that is being referred to by the word *bhujā* in both the citations as well as explanations offered below. Half of the side of the square (equal to the radius) is divided into n equal bits, known as *bhujā-khaṇḍas*. It is these *bhujā-khaṇḍas* $\left(\frac{r}{n}\right)$, $2\left(\frac{r}{n}\right)\cdots$ whose powers are summed.

To start with, *Yuktibhāṣā* discusses just the basic summation of *bhujā-khaṇḍas* called *Mūla-saṅkalita*. We now cite the following from the translation of *Yuktibhāṣā*:[68]

> Now is described the methods of making the summations (referred to in the earlier sections). At first, the simple arithmetical progression (*kevala-saṅkalita*) is described. This is followed by the summation of the products of equal numbers (squares). ...
> Here, in this *mūla-saṅkalita* (basic arithmetical progression), the final *bhujā* is

[68] *Gaṇita-yukti-bhāṣā*, cited above, Section 6.4, pp. 61–67, 192–97, 382–88.

equal to the radius. The term before that will be one segment (*khaṇḍa*) less. The next one will be two segments less. Here, if all the terms (*bhujās*) had been equal to the radius, the result of the summation would be obtained by multiplying the radius by the number of *bhujās*. However, here, only one *bhujā* is equal to the radius. And, from that *bhujā*, those associated with the smaller hypotenuses are less by one segment each, in order. Now, suppose the radius to be of the same number of units as the number of segments to which it has been divided, in order to facilitate remembering (their number). Then, the number associated with the penultimate *bhujā* will be less by one (from the number of units in the radius); the number of the next one, will be less by two from the number of units in the radius. This reduction (in the number of segments) will increase by one (at each step). The last reduction will practically be equal to the measure of the radius, for it will be less only by one segment. In other words, when the reductions are all added, the sum thereof will practically (*prāyeṇa*) be equal to the summation of the series from 1 to the number of units in the radius; it will be less only by one radius length. Hence, the summation will be equal to the product of the number of units in the radius with the number of segments plus one, and divided by 2. The summation of all the *bhujās* of the different hypotenuses is called *bhujā-saṅkalita*.
Now, the smaller the segments, the more accurate (*sūkṣma*) will be the result. Hence, do the summation also by taking each segment as small as an atom (*aṇu*). Here, if it (namely, the *bhujā* or the radius) is divided into *parārdha* (a very large number) parts, to the *bhujā* obtained by multiplying by *parārdha* add one part in *parārdha* and multiply by the radius and divide by 2, and then divide by *parārdha*. For, the result will practically be the square of the radius divided by two. ...

The first summation, the *bhujā-saṅkalita*, may be written in the reverse order from the final *bhujā* to the first *bhujā* as

$$S_n^{(1)} = \left(\frac{nr}{n}\right) + \left(\frac{(n-1)r}{n}\right) + ... + \left(\frac{r}{n}\right). \tag{85}$$

Now, conceive of the *bhujā-khaṇḍa* $\frac{r}{n}$ as being infinitesimal (*aṇu*) and at the same time as of unit-measure (*rūpa*), so that the radius will be the measure of n, the *pada*, or the number of terms. Then

$$S_n^{(1)} = n + (n-1) + ... + 1. \tag{86}$$

If each of the terms were of the measure of radius (n) then the sum would be nothing but n^2, the square of the radius. But only the first term is of the measure of radius, the next is deficient by one segment (*khaṇḍa*), the next by two segments and so on till the last term which is deficient by an amount equal to radius-minus-one segment. In other words,

$$\begin{aligned} S_n^{(1)} &= n + [n-1] + [n-2] .... + [n-(n-2)] + [n-(n-1)] \\ &= n.n - [1 + 2 + ... + (n-1)]. \end{aligned} \tag{87}$$

When n is very large, the quantity to be subtracted from n^2 is practically (*prāyeṇa*) the same as $S_n^{(1)}$, thus leading to the estimate

$$S_n^{(1)} \approx n^2 - S_n^{(1)}, \qquad (88)$$

$$\text{or} \quad S_n^{(1)} \approx \frac{n^2}{2}. \qquad (89)$$

It is stated that the result is more accurate, when the size of the segments are small (or equivalently, the value of n is large).[69]

If instead of making the approximation as in (88), we proceed with (87) as it is, we get $S_n^{(1)} = n^2 - (S_n^{(1)} - n)$, which leads to the well-known exact value of the sum of the first n natural numbers

$$S_n^{(1)} = \frac{n(n+1)}{2}, \qquad (90)$$

With the convention that the $\frac{r}{n}$ is of unit-measure, the above estimate (89) is stated in the form that the *bhujā-saṅkalita* is half the square of the radius.

12.2. Summation of squares (*Varga-saṅkalita*)

We now cite the following from the translation of *Yuktibhāṣā*:[70]

> Now is explained the summation of squares (*varga-saṅkalita*). Obviously, the squares of the *bhujās*, which are summed up above, are the *bhujās* each multiplied by itself. Here, if the *bhujās* which are all multipliers, had all been equal to the radius, their sum, (*saṅkalita* derived above), multiplied by the radius would have been the summation of their squares. Here, however, only one multiplier happens to be equal to the radius, and that is the last one. The one before that will have the number of segments one less than in the radius. (Hence) if that, (i.e., the second one), is multiplied by the radius, it would mean that one multiplied by the penultimate *bhujā* would have been the increase in the summation of the squares. Then (the segment) next below is the third. That will be less than the radius by two segments. If that is multiplied by the radius, it will mean that, the summation of the squares will increase by the product of the *bhujā* by two (segments). In this manner, the summation in which the multiplication is done by the radius (instead of the *bhujās*) would be larger than the summation of squares by terms

[69] Śaṅkara Vāriyar also emphasizes the same idea, in his discussion of the estimation of *saṅkalitas* in his commentary *Kriyākramakarī* on *Līlāvatī* (cited above (fn. 14), comm. on verse 199, p. 382.):

> खण्डस्याल्पत्वे सत्येव लब्धस्य सूक्ष्मता च स्यात्।
> Only when the segment is small (*khaṇḍasyālpatve*) the result obtained would be accurate.

[70] *Gaṇita-yukti-bhāṣā*, cited above, Section 6.4, pp. 61–67, 192–97, 382–88.

which involve the successively smaller *bhujās* multiplied by successively higher numbers. If (all these additions) are duly subtracted from the summation where the radius is used as the multiplier, the summation of squares (*varga-saṅkalita*) will result.

Now, the *bhujā* next to the east-west line is less than the radius by one (segment). So if all the excesses are summed up and added, it would be the summation of the basic summation (*mūla-saṅkalita-saṅkalita*). Because, the sums of the summations is verily the 'summation of summations' (*saṅkalita-saṅkalita*). There, the last sum has (the summation of) all the *bhujās*. The penultimate sum is next lower summation to the last. This penultimate sum is the summation of all the *bhujās* except the last *bhujā*. Next to it is the third sum which is the sum of all the *bhujās* except the last two. Thus, each sum of the *bhujās* commencing from any *bhujā* which is taken to be the last one in the series, will be less by one *bhujā* from the sum (of the *bhujās*) before that.

Thus, the longest *bhujā* is included only in one sum. But the *bhujā* next lower than the last (*bhujā*) is included both in the last sum and also in the next lower sum. The *bhujās* below that are included in the three, four etc. sums below it. Hence, it would result that the successively smaller *bhujās* commencing from the one next to the last, which have been multiplied by numbers commencing from 1 and added together, would be summation of summations (*saṅkalita-saṅkalita*). Now, it has been stated earlier that the summation (*saṅkalita*) of (the segments constituting) a *bhujā* which has been very minutely divided, will be equal to half the square of the last *bhujā*. Hence, it follows that, in order to obtain the summation (*saṅkalita*) of the *bhujās* ending in any particular *bhujā*, we will have to square each of the *bhujās* and halve it. Thus, the summation of summations (*saṅkalita-saṅkalita*) would be half the summation of the squares of all the *bhujās*. In other words, half the summation of the squares is the summation of the basic summation. So, when the summation is multiplied by the radius, it would be one and a half times the summation of the squares. This fact can be expressed by stating that this contains half more of the summation of squares. Therefore, when the square of the radius divided by two is multiplied by the radius and one-third of it subtracted from it, the remainder will be one-third of the whole. Thus it follows that one-third of the cube of the radius will be the summation of squares (*varga-saṅkalita*).

With the same convention that $\frac{r}{n}$ is the measure of the unit, the *bhujā-varga-saṅkalita* (the sum of the squares of the *bhujās*) will be

$$S_n^{(2)} = n^2 + (n-1)^2 + ... + 1^2. \tag{91}$$

In above expression, each *bhujā* is multiplied by itself. If instead, we consider that each *bhujā* is multiplied by the radius (n in our units), then that would give raise to the sum

$$n\,[n + (n-1) + ... + 1] = n\ S_n^{(1)}. \tag{92}$$

This sum is exceeds the *bhujā-varga-saṅkalita* by the amount

$$nS_n^{(1)} - S_n^{(2)} = 1.(n-1) + 2.(n-2) + 3.(n-3) + \ldots + (n-1).1.$$

This may be written in the form

$$\begin{aligned} nS_n^{(1)} - S_n^{(2)} = (n-1) + (n-2) + (n-3) &\quad + \ldots \quad +1 \\ +(n-2) + (n-3) &\quad + \ldots \quad +1 \\ +(n-3) &\quad + \ldots \quad +1 \\ &\quad + \ldots \quad . \end{aligned} \tag{93}$$

Thus,

$$nS_n^{(1)} - S_n^{(2)} = S_{n-1}^{(1)} + S_{n-2}^{(1)} + S_{n-3}^{(1)} + \ldots . \tag{94}$$

The right hand side of (94) is called the *saṅkalita-saṅkalita* (or *saṅkalitaikya*), the repeated sum of the sums $S_i^{(1)}$ (here taken in the order $i = n-1,\ n-2, \ldots 1$). These are defined also by Śaṅkara Vāriyar in *Kriyākramakarī* as follows:[71]

तथा हि सङ्कलितानां योगो हि सङ्कलितसङ्कलितमुच्यते। तत्र अन्त्यसङ्कलितं सर्वासां भुजानां योगः। उपान्त्यसङ्कलितं तु अन्त्यभुजाव्यतिरिक्तानामितरेषां योगः। उपान्त्यात् पूर्वस्य सङ्कलितं पुनस्तदवधिकानामेव भुजानां योगः। एवं पूर्व-सङ्कलितानि स्वोत्तरात् सङ्कलितात् एकैकेन भुजेन विरहितानि भवन्ति।

The sum of the summations is called as *saṅkalita-saṅkalita.* Of them the last *saṅkalita* is the sum all the *bhujās.* The penultimate *saṅkalita* is the sum of all the *bhujās* other than the last one. The *saṅkalita* of the one preceding the penultimate is the sum of the *bhujās* ending with that. Thus, all the preceding *saṅkalitas* will fall short by a *bhujā* from the succeeding *saṅkalita.*

For large n, we have already estimated in (89) that $S_n^{(1)} \approx \frac{n^2}{2}$. Thus, for large n

$$nS_n^{(1)} - S_n^{(2)} \approx \frac{(n-1)^2}{2} + \frac{(n-2)^2}{2} + \frac{(n-3)^2}{2} + \ldots . \tag{95}$$

Thus, the right hand side of (94) (the *saṅkalita-saṅkalita* or the excess of $nS_n^{(1)}$ over $S_n^{(2)}$) is essentially $\frac{S_n^{(2)}}{2}$ for large n, so that we obtain

$$nS_n^{(1)} - S_n^{(2)} \approx \frac{S_n^{(2)}}{2}. \tag{96}$$

[71] *Kriyākramakarī* on *Līlāvatī*, cited above (fn. 14), comm. on verse 199, pp. 382–83.

Again, using the earlier estimate (89) for $S_n^{(1)}$, we obtain the result

$$S_n^{(2)} \approx \frac{n^3}{3}. \qquad (97)$$

Thus *bhujā-varga-saṅkalita* is one-third the cube of the radius.

12.3. *Sama-ghāta-saṅkalita*

We now cite the following from the translation of *Yuktibhāṣā*:[72]

> Now, the square of the square (of a number) is multiplied by itself, it is called *sama-pañca-ghāta* (number multiplied by itself five times). The successive higher order summations are called *sama-pañcādi-ghāta-saṅkalita* (and will be the summations of powers of five and above). Among them if the summation (*saṅkalita*) of powers of some order is multiplied by the radius, then the product is the summation of summations (*saṅkalita-saṅkalita*) of the (powers of the) multiplicand (of the given order), together with the summation of powers (*sama-ghāta-saṅkalita*) of the next order. Hence, to derive the summation of the successive higher powers: Multiply each summation by the radius. Divide it by the next higher number and subtract the result from the summation got before. The result will be the required summation to the higher order.
>
> Thus, divide by two the square of the radius. If it is the cube of the radius, divide by three. If it is the radius raised to the power of four, divide by four. If it is (the radius) raised to the power of five, divide by five. In this manner, for powers rising one by one, divide by numbers increasing one by one. The results will be, in order, the summations of powers of numbers (*sama-ghāta-saṅkalita*). Here, the basic summation is obtained from the square, the summation of squares from the cube, the summation of cubes from the square of the square. In this manner, if the numbers are multiplied by themselves a certain number of times (i.e., raised to a certain degree) and divided by the same number, that will be the summation of the order one below that. Thus (has been stated) the method of deriving the summations of (natural) numbers, (their) squares etc.

In the case of a general *samaghāta-saṅkalita*, (summation of equal powers) given by

$$S_n^{(k)} = n^k + (n-1)^k + \ldots + 1^k, \qquad (98)$$

the procedure followed to estimate its behavior for large n is essentially the same as that followed in the case of *vargasaṅkalita*. We first compute the excess of $nS_n^{(k-1)}$ over $S_n^{(k)}$ to be a *saṅkalita-saṅkalita* or repeated sum of the lower order

[72] *Gaṇita-yukti-bhāṣā*, cited above, Section 6.4, pp. 61–67, 192–97, 382–88.

saṅkalitas $S_r^{(k-1)}$:

$$nS_n^{(k-1)} - S_n^{(k)} = S_{n-1}^{(k-1)} + S_{n-2}^{(k-1)} + S_{n-3}^{(k-1)} + \dots \quad (99)$$

If the lower order *saṅkalita* $S_n^{(k-1)}$ has already been estimated to be, say,

$$S_n^{(k-1)} \approx \frac{n^k}{k}, \quad (100)$$

then, the above relation (99) leads to[73]

$$\begin{aligned} nS_n^{(k-1)} - S_n^{(k)} &\approx \frac{(n-1)^k}{k} + \frac{(n-2)^k}{k} + \frac{(n-3)^k}{k} + \dots \\ &\approx \left(\frac{1}{k}\right) S_n^{(k)}. \end{aligned} \quad (101)$$

Rewriting the above equation we have[74]

$$S_n^{(k)} \approx nS_n^{(k-1)} - \left(\frac{1}{k}\right) S_n^{(k)}. \quad (102)$$

Using (100), we obtain the estimate

$$S_n^{(k)} \approx \frac{n^{k+1}}{(k+1)}. \quad (103)$$

[73]As one of the reviewers has pointed out, this argument leading to (101) is indeed similar to the derivation of the following relation, which is based on the interchange of order in iterated integrals:

$$\int_0^1 (1-x)x^{k-1}dx \quad = \quad \int_0^1 x^{k-1} \int_x^1 dy\, dx = \int_0^1 y \int_0^y x^{k-1} dx\, dy = \int_0^1 \frac{y^k}{k} dy.$$

[74]As Śaṅkara Vāriyar states in his *Kriyākramakarī* on *Līlāvatī* (cited above (fn. 14), p. 383):

अत उत्तरोत्तरसङ्कलितानयनाय तत्तत्सङ्कलितस्य व्यासार्धगुणनम्
एकैकाधिकसङ्ख्याप्तस्वांशशोधनं च कार्यम् इति स्थितम्।

Therefore it is established that, for obtaining the sum of the next order, the previous sum, has to be multiplied by the radius and the present sum, divided by one more than the previous [order], has to be diminished [from that product].

12.4. Repeated summations (*Saṅkalita-saṅkalita*)

After having estimated the sum of powers of natural numbers *samaghāta-saṅkalita* *Yuktibhāṣā* goes on to derive an estimate for the repeated summation (*saṅkalita-saṅkalita* or *saṅkalitaikya* or *vārasaṅkalita*) of the natural number $1, 2, \cdots, n$.[75]

> Now, are explained the first, second and further summations: The first summation (*ādya-saṅkalita*) is the basic summation (*mūla-saṅkalita*) itself. It has already been stated (that this is) half the product of the square of the number of terms (*pada-vargārdha*). The second (*dvitīya-saṅkalita*) is the summation of the basic summation (*mūla-saṅkalitaikya*). It has been stated earlier that it is equal to half the summation of squares. And that will be one-sixth of the cube of the number of terms.
>
> Now, the third summation: For this, take the second summation as the last term (*antya*); subtract one from the number of terms, and calculate the summation of summations as before. Treat this as the penultimate. Then subtract two from the number of terms and calculate the summation of summations. That will be the next lower term. In order to calculate the summation of summations of numbers in the descending order, the sums of one-sixths of the cubes of numbers in descending order would have to be calculated. That will be the summation of one-sixth of the cubes. And that will be one-sixth of the summation of cubes. As has been enunciated earlier, the summation of cubes is one-fourth the square of the square. Hence, one-sixth of one-fourth the square of the square will be the summation of one-sixth of the cubes. Hence, one-twenty-fourth of the square of the square will be the summation of one-sixth of the cubes. Then, the fourth summation will be, according to the above principle, the summation of one-twenty-fourths of the square of squares. This will also be equal to one-twenty-fourth of one-fifth of the fifth power. Hence, when the number of terms has been multiplied by itself a certain number of times, (i.e., raised to a certain degree), and divided by the product of one, two, three etc. up to that index number, the result will be the summation up to that index number amongst the first, second etc. summations (*ādya-dvitīyādi-saṅkalita*).

The first summation (*ādya-saṅkalita*) $V_n^{(1)}$ is just the *mūla-saṅkalita* or the basic summation of natural numbers, which has already been estimated in (89)

$$V_n^{(1)} = S_n^{(1)} = n + (n-1) + (n-2) + \ldots + 1 \approx \frac{n^2}{2}. \tag{104}$$

The second summation (*dvitīya-saṅkalita* or *saṅkalita-saṅkalita* or *saṅkalitaikya*) is given by

$$\begin{aligned} V_n^{(2)} &= V_n^{(1)} + V_{n-1}^{(1)} + V_{n-2}^{(1)} + \ldots \\ &= S_n^{(1)} + S_{n-1}^{(1)} + S_{n-2}^{(1)} + \ldots \end{aligned} \tag{105}$$

[75] *Gaṇita-yukti-bhāṣā*, cited above, Section 6.4, pp. 61–67, 192–97, 382–88.

As was done earlier, this second summation can be estimated using the estimate (89) for $S_n^{(1)}$

$$V_n^{(2)} \approx \frac{n^2}{2} + \frac{(n-1)^2}{2} + \frac{(n-2)^2}{2} + \dots \quad (106)$$

Therefore

$$V_n^{(2)} \approx \left(\frac{1}{2}\right) S_n^{(2)}. \quad (107)$$

Using the earlier estimate (97) for $S_n^{(2)}$, we get an estimate for the *dvitīya-saṅkalita*

$$V_n^{(2)} \approx \frac{n^3}{6}. \quad (108)$$

Now the next repeated summation can be found in the same way

$$\begin{aligned} V_n^{(3)} &= V_n^{(2)} + V_{n-1}^{(2)} + V_{n-2}^{(2)} + \dots \\ &\approx \frac{n^3}{6} + \frac{(n-1)^3}{6} + \frac{(n-2)^3}{6} + \dots \\ &\approx \left(\frac{1}{6}\right) S_n^{(3)} \\ &\approx \frac{n^4}{24}. \end{aligned} \quad (109)$$

It is noted that proceeding this way we can estimate repeated summation $V_n^{(k)}$ of order k, for large n, to be[76]

$$\begin{aligned} V_n^{(k)} &= V_n^{(k-1)} + V_{n-1}^{(k-1)} + V_{n-2}^{(k-1)} + \dots \\ &\approx \frac{n^{k+1}}{1.2.3.\dots(k+1)}. \end{aligned} \quad (110)$$

[76] These are again estimates for large n. As mentioned in Section 4, exact expressions for the first two summations, $V_n^{(1)}$ and $V_n^{(2)}$, are given in *Āryabhaṭīya, Gaṇitapāda* 21; and the exact expression for the k-th order repeated summation $V_n^{(k)}$ has been given (under the name *vāra-saṅkalita*), by Nārāyaṇa Paṇḍita (c. 1350) in his *Gaṇitakaumudī*, 3.19. This exact expression for $V_n^{(k)}$ is also noted in Section 7.5.3 of *Yuktibhāṣā*.

13. Derivation of the Mādhava series for π

The following accurate value of π (correct to 11 decimal places), given by Mādhava, has been cited by Nīlakaṇṭha in his *Āryabhaṭīya-bhāṣya* and by Śaṅkara Vāriyar in his *Kriyākramakarī*.[77]

विबुधनेत्रगजाहिहुताशनत्रिगुणवेदभवारणबाहवः ।
नवनिखर्वमिते वृतिविस्तरे परिधिमानमिदं जगदुर्बुधाः ॥

The π value given above is:

$$\pi \approx \frac{2827433388233}{9 \times 10^{11}} = 3.141592653592... \quad (111)$$

The 13 digit number appearing in the numerator has been specified using *bhūta-saṅkhya* system, whereas the denominator is specified by word numerals.[78]

13.1. Infinite series for π

The infinite series for π attributed to Mādhava is cited by Śaṅkara Vāriyar in his commentaries *Kriyākramakarī* and *Yukti-dīpikā*. Mādhava's verse quoted runs as follows:[79]

व्यासे वारिधिनिहते रूपहृते व्याससागराभिहते।
त्रिशरादिविषमसङ्ख्याभक्तमृणं स्वं पृथक् क्रमात् कुर्यात्॥

The diameter multiplied by four and divided by unity [is found and saved]. Again the products of the diameter and four are divided by the odd numbers like three, five, etc., and the results are subtracted and added in order [to the earlier result saved].

The series given by the verse may be represented as

$$Paridhi = 4 \times Vy\bar{a}sa \times \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots\ldots\right). \quad (112)$$

---

[77] *Āryabhaṭīya-bhāṣya* on *Āryabhaṭīya*, cited above (fn. 53), comm. on *Gaṇitapāda* 10, p. 42; *Kriyākramakarī* on *Līlāvatī*, cited above (fn. 14), comm. on verse 199, p. 377.

[78] In the *bhūta-saṅkhyā* system, *vibudha* =33, *netra* =2, *gaja* =8, *ahi* =8, *hutāśana* =3, *triguṇa* =3, *veda* =4, *bha* =27, *vāraṇa* =8, *bāhu* =2. In word numerals, *nikharva* represents 10^{11}. Hence, *nava-nikharva* $= 9 \times 10^{11}$.

[79] op. cit., p. 379.

The words *paridhi* and *vyāsa*[80] in the above equation refer to the circumference and diameter respectively. Hence the equation may be rewritten as

$$\frac{\pi}{4} = \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots \qquad \right). \tag{113}$$

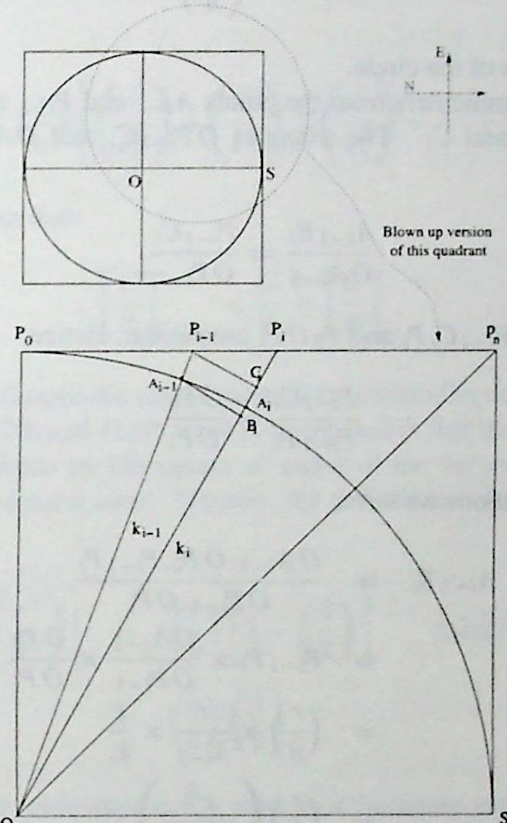


FIGURE 8. Geometrical construction used in the proof of the infinite series for π.

We shall now present the derivation of the above result as outlined in *Yuktibhāṣā* of Jyeṣṭhadeva and *Kriyākramakarī* of Śaṅkara Vāriyar. For this purpose, let us consider the quadrant $O P_0 P_n S$ of the square circumscribing the given circle (see Figure 8). Divide the side $P_0 P_n$ into n equal parts (n very large). $P_0 P_i$'s

---

[80] Nīlakaṇṭha, in his *Āryabhaṭīya-bhāṣya*, presents the etymological derivation of the word *vyāsa* as 'the one which splits the circle into two halves': व्यासेन हि वृत्तं व्यस्यते। (*Āryabhaṭīya-bhāṣya*, cited above (fn. 53), comm. on *Gaṇitapāda* 11, p. 43).

are the *bhujās* and OP_i's are the *karṇas* denoted by k_i. The points of intersection of these *karṇas* and the circle are marked as A_is.

The *bhujās* P_0P_i, the *karṇas* k_i and the east-west line OP_0 form right-angled triangles whose hypotenuses are given by

$$k_i^2 = r^2 + \left(\frac{ir}{n}\right)^2, \tag{114}$$

where r is the radius of the circle.

The feet of perpendiculars from the points A_{i-1} and P_{i-1} along the i^{th} *karṇa* are denoted by B_i and C_i. The triangles $OP_{i-1}C_i$ and $OA_{i-1}B_i$ are similar. Hence,

$$\frac{A_{i-1}B_i}{OA_{i-1}} = \frac{P_{i-1}C_i}{OP_{i-1}}. \tag{115}$$

Similarly triangles $P_{i-1}C_iP_i$ and P_0OP_i are similar. Hence,

$$\frac{P_{i-1}C_i}{P_{i-1}P_i} = \frac{OP_0}{OP_i}. \tag{116}$$

From these two relations we have,

$$\begin{aligned} A_{i-1}B_i &= \frac{OA_{i-1}.OP_0.P_{i-1}P_i}{OP_{i-1}.OP_i} \\ &= P_{i-1}P_i \times \frac{OA_{i-1}}{OP_{i-1}} \times \frac{OP_0}{OP_i} \\ &= \left(\frac{r}{n}\right) \times \frac{r}{k_{i-1}} \times \frac{r}{k_i} \\ &= \left(\frac{r}{n}\right)\left(\frac{r^2}{k_{i-1}k_i}\right). \end{aligned} \tag{117}$$

It is then noted that when n is large, the Rsines $A_{i-1}B_i$ can be taken as the arc-bits themselves.

परिधिखण्डस्यार्धज्या $\rightarrow$ परिध्यंश

i.e., $A_{i-1}B_i \rightarrow \widehat{A_{i-1}A_i}$.

Thus, $\frac{1}{8}$th of the circumference of the circle can be written as sum of the contributions given by (117). That is

$$\frac{C}{8} \approx \left(\frac{r}{n}\right)\left[\left(\frac{r^2}{k_0k_1}\right)+\left(\frac{r^2}{k_1k_2}\right)+\left(\frac{r^2}{k_2k_3}\right)+\cdots+\left(\frac{r^2}{k_{n-1}k_n}\right)\right]. \tag{118}$$

Though this is the expression that actually needs to be evaluated, the text mentions that there may not be much difference in approximating it by either of the following expressions:

$$\left[\frac{C}{8}\right]_{left} = \left(\frac{r}{n}\right)\left[\left(\frac{r^2}{k_0^2}\right)+\left(\frac{r^2}{k_1^2}\right)+\left(\frac{r^2}{k_2^2}\right)+\cdots+\left(\frac{r^2}{k_{n-1}^2}\right)\right] \tag{119}$$

or

$$\left[\frac{C}{8}\right]_{right} = \left(\frac{r}{n}\right)\left[\left(\frac{r^2}{k_1^2}\right)+\left(\frac{r^2}{k_2^2}\right)+\left(\frac{r^2}{k_3^2}\right)+\cdots+\left(\frac{r^2}{k_n^2}\right)\right]. \tag{120}$$

It can be easily seen that

$$\left[\frac{C}{8}\right]_{right} < \frac{C}{8} < \left[\frac{C}{8}\right]_{left}. \tag{121}$$

In other words, though the actual value of the circumference lies in between the values given by (120) and (119) what is being said is that there will not be much difference if we divide by the square of either of the *karṇas* rather than by the product of two successive ones. Actually, the difference between (120) and (119) is given by

$$\begin{aligned}\left(\frac{r}{n}\right)\left[\left(\frac{r^2}{k_0^2}\right)-\left(\frac{r^2}{k_n^2}\right)\right] &= \left(\frac{r}{n}\right)\left[1-\left(\frac{1}{2}\right)\right] \qquad (\text{since } k_0^2, k_n^2 = r^2, 2r^2)\\ &= \left(\frac{r}{n}\right)\left(\frac{1}{2}\right) \end{aligned} \tag{122}$$

Evidently this difference approaches zero as n becomes very large, as noted in both the texts *Yuktibhāṣā* and *Kriyākramakarī*.

The terms in (120) are evaluated using the *śodhya-phala* technique (binomial series, discussed earlier in Section 11) and each one of them may be re-written in the form[81]

$$\frac{r}{n}\left(\frac{r^2}{k_i^2}\right) = \frac{r}{n} - \frac{r}{n}\left(\frac{k_i^2-r^2}{r^2}\right) + \frac{r}{n}\left(\frac{k_i^2-r^2}{r^2}\right)^2 - \ldots \tag{123}$$

[81] It may be noted that this series is convergent since $k_i^2 = r^2 + \left(\frac{ir}{n}\right)^2$ and $0 \le (k_i^2 - r^2) < r^2$ for $i < n$.

Using (114) and (123) in (120), we obtain:

$$\begin{aligned}
\frac{C}{8} &= \sum_{i=1}^{n} \frac{r}{n}\left(\frac{r^2}{k_i^2}\right) \\
&= \sum_{i=1}^{n} \left(\frac{r}{n}\right)\left(\frac{r^2}{r^2+\left(\frac{ir}{n}\right)^2}\right) &(124)\\
&= \sum_{i=1}^{n} \left[\frac{r}{n} - \frac{r}{n}\left(\frac{\left(\frac{ir}{n}\right)^2}{r^2}\right) + \frac{r}{n}\left(\frac{\left(\frac{ir}{n}\right)^2}{r^2}\right)^2 - \ldots\right] &(125)\\
&= \left(\frac{r}{n}\right)[1+1+\ldots+1] \\
&\quad - \left(\frac{r}{n}\right)\left(\frac{1}{r^2}\right)\left[\left(\frac{r}{n}\right)^2 + \left(\frac{2r}{n}\right)^2 + \ldots + \left(\frac{nr}{n}\right)^2\right] \\
&\quad + \left(\frac{r}{n}\right)\left(\frac{1}{r^4}\right)\left[\left(\frac{r}{n}\right)^4 + \left(\frac{2r}{n}\right)^4 + \ldots + \left(\frac{nr}{n}\right)^4\right] \\
&\quad - \left(\frac{r}{n}\right)\left(\frac{1}{r^6}\right)\left[\left(\frac{r}{n}\right)^6 + \left(\frac{2r}{n}\right)^6 + \ldots + \left(\frac{nr}{n}\right)^6\right] \\
&\quad + \ldots\ . &(126)
\end{aligned}$$

Each of the terms in (126) is a sum of results (*phala-yoga*) which we need to estimate when n is very large, and we have a series of them (*phala-paramparā*) which are alternatively positive and negative. Clearly the first term is just the sum of the *bhujā-khaṇḍas*.

The *bhujās* themselves are given by the integral multiples of *bhujā-khaṇḍa*, namely, $\frac{r}{n}, \frac{2r}{n}, \ldots \frac{nr}{n}$. In the series expression for the circumference given above, we thus have the *saṅkalitas* or summations of even powers of the *bhujās*, such as the *bhujā-varga-saṅkalita*, $\left(\frac{r}{n}\right)^2 + \left(\frac{2r}{n}\right)^2 + \ldots. + \left(\frac{nr}{n}\right)^2$, *bhujā-varga-varga-saṅkalita*, $\left(\frac{r}{n}\right)^4 + \left(\frac{2r}{n}\right)^4 + \ldots.. + \left(\frac{nr}{n}\right)^4$, and so on.

If we take out the powers of *bhujā-khaṇḍa* $\frac{r}{n}$, the summations involved are those of even powers of the natural numbers, namely *edādyekottara-varga-saṅkalita*, $1^2 + 2^2 + \ldots + n^2$, *edādyekottara-varga-varga-saṅkalita*, $1^4 + 2^4 + \ldots + n^4$, and so on.

Now, recalling the estimates that were obtained earlier for these *saṅkalitas*, when n is large,

$$\sum_{i=1}^{n} i^k \approx \frac{n^{k+1}}{k+1}, \tag{127}$$

we arrive at the result[82]

$$\frac{C}{8} = r\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right), \tag{128}$$

which is same as (112).

14. Derivation of end-correction terms (*Antya-saṃskāra*)

It is well known that the series given by (112) for $\frac{\pi}{4}$ is an extremely slowly converging series. It is so slow that even for obtaining the value of π correct to 2 decimal places one has to find the sum of hundreds of terms and for getting it correct to 4-5 decimal places we need to consider millions of terms. Mādhava seems to have found an ingenious way to circumvent this problem. The technique employed by Mādhava is known as *antya-saṃskāra*. The nomenclature stems from the fact that a correction (*saṃskara*) is applied towards the end (*anta*) of the series, when it is terminated after considering only a certain number of terms from the beginning.

14.1. The criterion for *antya-saṃskāra* to yield accurate result

The discussion on *antya-saṃskāra* in both *Yuktibhāṣā* and *Kriyākramakarī* commences with the question:

> How is it that one obtains the value of the circumference more accurately by doing *antya-saṃskara*, instead of repeatedly dividing by odd numbers? [83]

[82] In modern terminology, the above derivation amounts to the evaluation of the following integral

$$\frac{C}{8} = \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{r}{n}\right)\left(\frac{r^2}{r^2+\left(\frac{ir}{n}\right)^2}\right) = r\int_0^1 \frac{dx}{1+x^2}.$$

[83] कथं पुनरत्र मुहुर्विषमसङ्ख्याहरणेन लभ्यस्य परिधेः आसन्नत्वम् अन्त्यसंस्कारेण आपाद्यते । उच्यते ।... (*Kriyākramakarī* on *Līlāvatī*, cited above (fn. 14), comm. on verse 199, p. 386.)

The argument adduced in favor of terminating the series at any desired term, still ensuring the accuracy, is as follows. Let the series for $\frac{\pi}{4}$ be written as

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \ldots + (-1)^{\frac{p-3}{2}} \frac{1}{p-2} + (-1)^{\frac{p-1}{2}} \frac{1}{a_{p-2}}, \tag{129}$$

where $\frac{1}{a_{p-2}}$ is the correction term applied after odd denominator $p - 2$. On the other hand, if the correction term $\frac{l}{a_p}$, is applied after the odd denominator p, then

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \ldots + (-1)^{\frac{p-3}{2}} \frac{1}{p-2} + (-1)^{\frac{p-1}{2}} \frac{1}{p} + (-1)^{\frac{p+1}{2}} \frac{1}{a_p}. \tag{130}$$

If the correction terms indeed lead to the exact result, then both the series (129) and (130) should yield the same result. That is,

$$\frac{1}{a_{p-2}} = \frac{1}{p} - \frac{1}{a_p} \quad \text{or} \quad \frac{1}{a_{p-2}} + \frac{1}{a_p} = \frac{1}{p}, \tag{131}$$

is the criterion that must be satisfied for the end-correction (*antya-saṃskāra*) to lead to the exact result.

14.2. Successive approximations to get more accurate correction-terms

The criterion given by (131) is trivially satisfied when we choose $a_{p-2} = a_p = 2p$. However, this value $2p$ cannot be assigned to both the correction-divisors[84] a_{p-2} and a_p because both the corrections should follow the same rule. That is,

$$\begin{aligned} a_{p-2} &= 2p, \quad \Rightarrow \quad a_p = 2(p+2) \\ \text{or,} \quad a_p &= 2p, \quad \Rightarrow \quad a_{p-2} = 2(p-2). \end{aligned}$$

We can, however, have both a_{p-2} and a_p close to $2p$ by taking $a_{p-2} = 2p - 2$ and $a_p = 2p + 2$, as there will always persist this much difference between $p - 2$ and p when they are doubled. Hence, the first (order) estimate of the correction divisor is given as, "double the even number above the last odd-number divisor p",

$$a_p = 2(p+1). \tag{132}$$

But, it can be seen right away that, with this value of the correction divisor, the condition for accuracy (131), stated above, is not exactly satisfied. Therefore a

[84] By the term correction-divisor (*saṃskāra-hāraka*) is meant the divisor of the correction term.

measure of inaccuracy (*sthaulya*) $E(p)$ is introduced

$$E(p) = \left[\frac{1}{a_{p-2}} + \frac{1}{a_p}\right] - \frac{1}{p} . \tag{133}$$

Now, since the error cannot be eliminated, the objective is to find the correction denominators a_p such that the inaccuracy $E(p)$ is minimised. When we set $a_p = 2(p+1)$, the inaccuracy will be

$$\begin{aligned} E(p) &= \left[\frac{1}{(2p-2)} + \frac{1}{(2p+2)}\right] - \frac{1}{p} \\ &= \frac{1}{(p^3 - p)} . \end{aligned} \tag{134}$$

This estimate of the inaccuracy, E_p being positive, shows that the correction has been over done and hence there has to be a reduction in the correction. This means that the correction-divisor has to be increased. If we take $a_p = 2p + 3$, thereby leading to $a_{p-2} = 2p - 1$, we have

$$\begin{aligned} E(p) &= \left[\frac{1}{(2p-1)} + \frac{1}{(2p+3)}\right] - \frac{1}{p} \\ &= \frac{(-2p+3)}{(4p^3 + 4p^2 - 3p)} . \end{aligned} \tag{135}$$

Now, the inaccuracy happens to be negative. But, more importantly, it has a term proportional to p in the numerator. Hence, for large p, $E(p)$ given by (135) varies inversely as p^2, while for the divisor given by (132), $E(p)$ as given by (134) varied inversely as p^3.[85]

From (134) and (135) it is obvious that, if we want to reduce the inaccuracy and thereby obtain a better correction, then a number less than 1 has to be added to the correction-divisor (132) given above. If we try adding *rūpa* (unity) divided by the correction divisor itself, i.e., if we set $a_p = 2p+2+\frac{1}{(2p+2)}$, the contributions from the correction-divisors get multiplied essentially by $\left(\frac{1}{2p}\right)$. Hence, to get rid of the higher order contributions, we need an extra factor of 4, which will be achieved if we take the correction divisor to be

$$a_p = (2p+2) + \frac{4}{(2p+2)} = \frac{(2p+2)^2 + 4}{(2p+2)} . \tag{136}$$

[85] It may be noted that among all possible correction divisors of the type $a_p = 2p + m$, where m is an integer, the choice of $m = 2$ is optimal, as in all other cases there will arise a term proportional to p in the numerator of the inaccuracy $E(p)$.

Then, correspondingly, we have

$$a_{p-2} = (2p-2) + \frac{4}{(2p-2)} = \frac{(2p-2)^2+4}{(2p-2)}. \tag{137}$$

We can then calculate the inaccuracy to be

$$\begin{aligned} E(p) &= \left[\frac{1}{(2p-2)+\frac{4}{2p-2}} + \frac{1}{(2p+2)+\frac{4}{2p+2}} \right] - \left(\frac{1}{p}\right) \\ &= \left[\frac{(4p^3)}{(4p^4+16)} \right] - \frac{(16p^4+64)}{4p(4p^4+16)} \\ &= \frac{-4}{(p^5+4p)} \end{aligned} \tag{138}$$

Clearly, the *sthaulya* with this (second order) correction divisor has improved considerably, in that it is now proportional to the inverse fifth power of the odd number.[86]

At this stage, we may display the result obtained for the circumference with the correction term as follows. If only the first order correction (132) is employed, we have

$$C = 4d\left[1 - \frac{1}{3} + \ldots + (-1)^{\frac{(p-1)}{2}}\frac{1}{p} + (-1)^{\frac{(p+1)}{2}}\frac{1}{(2p+2)}\right]. \tag{139}$$

If the second order correction (136) is taken into account, we have

$$\begin{aligned} C &= 4d\left[1 - \frac{1}{3} + \ldots + (-1)^{\frac{(p-1)}{2}}\frac{1}{p} + (-1)^{\frac{(p+1)}{2}}\frac{1}{(2p+2)+\frac{4}{(2p+2)}}\right] \\ &= 4d\left[1 - \frac{1}{3} + \ldots + (-1)^{\frac{(p-1)}{2}}\frac{1}{p} + (-1)^{\frac{(p+1)}{2}}\frac{\frac{(p+1)}{2}}{(p+1)^2+1}\right]. \end{aligned} \tag{140}$$

---

[86] It may be noted that if we take any other correction-divisor $a_p = 2p + 2 + \frac{m}{(2p+2)}$, where m is an integer, we will end up having a contribution proportional to p^2 in the numerator of the inaccuracy $E(p)$, unless $m = 4$. Thus the above form (136) is the optimal second order choice for the correction-divisor.

The verse due to Mādhava that we cited earlier as defining the infinite series for $\frac{\pi}{4}$ is, in fact, the first of a group of four verses that present the series along with the above end-correction.[87]

व्यासे वारिधिनिहते रूपहृते व्याससागराभिहते।
त्रिशरादिविषमसङ्ख्याभक्तमृणं स्वं पृथक् क्रमात् कुर्यात्॥
यत्सङ्ख्ययाऽत्र हरणे कृते निवृत्ता हृतिस्तु जामितया।
तस्या ऊर्ध्वगताया समसङ्ख्या तद्दलं गुणोऽन्ते स्यात्॥
तद्वर्गो रूपयुतो हारो व्यासाब्धिघाततः प्राग्वत्।
ताभ्यामाप्तं स्वमृणे कृते धने क्षेप एव करणीयः॥
लब्धः परिधिः सूक्ष्मो बहुकृत्वो हरणतोऽतिसूक्ष्मः स्यात्॥

The diameter multiplied by four and divided by unity. Again the products of the diameter and four are divided by the odd numbers like three, five, etc., and the results are subtracted and added in order.

Take half of the succeeding even number as the multipler at whichever [odd] number the division process is stopped, because of boredom. The square of that [even number] added to unity is the divisor. Their ratio has to be multiplied by the product of the diameter and four as earlier.

The result obtained has to be added if the earlier term [in the series] has been subtracted and subtracted if the earlier term has been added. The resulting circumference is very accurate; in fact more accurate than the one which may be obtained by continuing the division process [with large number of terms in the series].

Continuing this process further, *Yuktibhāṣā* presents the next order correction-term which is said to be even more accurate:[88]

अन्ते समसङ्ख्यादलवर्गः सैको गुणः स एव पुनः ॥
युगगुणितो रूपयुतः समसङ्ख्यादलहतो भवेद् हारः ।

At the end, [i.e., after terminating the series at some point, apply the correction term with] the multiplier being square of half of the [next] even number plus 1, and the divisor being four times the same multiplier with 1 added and multiplied by half the even number.

[87] *Kriyākramakarī* on *Līlāvatī*, cited above (fn. 14), comm. on verse 199, p. 379.

[88] *Gaṇita-yukti-bhāṣā*, cited above, p. 82; Also cited in *Yukti-dīpikā* on *Tantrasaṅgraha*, cited above (fn. 49), comm. on verse 2.1, p. 103.

In other words,[89]

$$\frac{1}{a_p} = \frac{\left(\frac{p+1}{2}\right)^2 + 1}{[(p+1)^2 + 4 + 1]\left(\frac{p+1}{2}\right)} = \frac{1}{(2p+2) + \cfrac{4}{2p+2+\cfrac{16}{2p+2}}}. \tag{141}$$

Hence, a much better approximation for $\frac{\pi}{4}$ is:[90]

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + \frac{1}{p} - \frac{\left(\frac{p+1}{2}\right)^2 + 1}{[(p+1)^2 + 4 + 1]\left(\frac{p+1}{2}\right)}. \tag{142}$$

[89] The inaccuracy or *sthaulya* associated with this correction can be calculated to be

$$E(p) = \frac{2304}{(64p^7 + 448p^5 + 1792p^3 - 2304p)}.$$

The inaccuracy now is proportional to the inverse seventh power of the odd-number. Again it can be shown that the number 16 in (141) is optimally chosen, in that any other choice would introduce a term proportional to p^2 in the numerator of $E(p)$, given above.

In fact, it has been noted by C. T. Rajagopal and M. S. Rangachari that D. T. Whiteside has shown (personal communication of D. T. Whiteside cited in C. T. Rajagopal and M. S. Rangachari, 'On an untapped source of medieval Kerala mathematics', Arch. for Hist. Sc. 35(2), 89–102, 1978), that the end correction-term can be exactly represented by the following continued fraction

$$\frac{1}{a_p} = \frac{1}{(2p+2) + \cfrac{2^2}{(2p+2) + \cfrac{4^2}{(2p+2) + \cfrac{6^2}{(2p+2) + \ldots}}}}.$$

[90] It may be noted that this correction term leads to a value of π, which is accurate up to 11 decimal places, when we merely evaluate terms up to $n = 50$ in the series (142). Incidentally the value of π, given in the rule *vibudhanetra...*, attributed to Mādhava that was cited in the beginning of Section 13, is also accurate up to 11 decimal places.

15. Transforming the Mādhava series for better convergence

After the estimation of end-correction terms, *Yuktibhāṣā* goes on to outline a method of transforming the Mādhava series (by making use of the above end-correction terms) to obtain new series that have much better convergence properties. We now reproduce the following from the English translation of *Yuktibhāṣā*:[91]

> Therefore, the circumference (of a circle) can be derived in taking into consideration what has been stated above. A method for that is stated in the verse
>
> समपञ्चाहतयो या रूपाद्ययुजां चतुर्घ्नमूलयुताः ताभिः ।
> षोडशगुणितात् व्यासाद् पृथगाहतेषु विषमयुतेः ।
> समफलयुतिमपहाय स्यादिष्टव्याससंभवः परिधिः ॥(I)
>
> The fifth powers of the odd numbers (1, 3, 5 etc.) are increased by 4 times themselves. The diameter is multiplied by 16 and it is successively divided by the (series of) numbers obtained (as above). The odd (first, third etc.) quotients obtained are added and are subtracted from the sum of the even (the second, fourth etc.) quotients. The result is the circumference corresponding to the given diameter.
>
> Herein above is stated a method for deriving the circumference. If the correction term is applied to an approximate circumference and the amount of inaccuracy (*sthaulya*) is found, and if it is additive, then the result is higher. Then it will become more accurate when the correction term obtained from the next higher odd number is subtracted. Since it happens that (an approximate circumference) becomes more and more accurate by making corrections in succeeding terms, if the corrections are applied right from the beginning itself, then the circumference will come out accurate. This is the rationale for this (above-stated result).
>
> When it is presumed that the correction-divisor is just double the odd number, the following is a method to obtain the (accurate) circumference by a correction for the corresponding inaccuracy (*sthaulyāṃśa-parihāra*), which is given by the verse:
>
> व्यासाद् वारिधिनिहतात् पृथगाप्तं त्र्याद्ययुग्विमूलघनैः ।
> त्रिघ्नव्यासे स्वमृणं क्रमशः कृत्वा परिधिरानेयः ॥(II)
>
> The diameter is multiplied by 4 and is divided, successively, by the cubes of the odd numbers beginning from 3, which are diminished by these numbers themselves. The diameter is now multiplied by three, and the quotients obtained above, are added to or subtracted from, alternatively. The circumference is to be obtained thus.
>
> If, however, it is taken that half the result (of dividing) by the last even number is taken as the correction, there is a method to derive the circumference by that way also, as given by the verse
>
> द्व्यादियुजां वा कृतयोः व्येका हाराद् द्विनिघ्नविष्कम्भे ।
> धनम् ऋणमन्तेऽन्त्योर्ध्वगतौजकृतिर्द्विसहिता हरस्यार्धम् ॥(III)
>
> The squares of even numbers commencing from 2, diminished by

[91] *Gaṇita-yukti-bhāṣā*, cited above, Section 6.9, pp. 80–82, 205–07, 402–04.

> one, are the divisors for four times the diameter. (Make the several divisions). The quotients got by (the division) are alternately added to or subtracted from twice the diameter. In the end, divide four times the diameter by twice the result of squaring the odd number following the last even number to which is added 2.

The method of *sthaulya-parihāra*, outlined above, essentially involves incorporating the correction terms into the series from the beginning itself. Let us recall that inaccuracy or *sthaulya* at each stage is given by

$$E(p) = \frac{1}{a_{p-2}} + \frac{1}{a_p} - \left(\frac{1}{p}\right). \tag{143}$$

The series for the circumference (112) can be expressed in terms of these *sthaulyas* as follows:

$$\begin{aligned} C &= 4d\left[\left(1-\frac{1}{a_1}\right)+\left(\frac{1}{a_1}+\frac{1}{a_3}-\frac{1}{3}\right)-\left(\frac{1}{a_3}+\frac{1}{a_5}-\frac{1}{5}\right)-\ldots\right] \\ &= 4d\left[\left(1-\frac{1}{a_1}\right)+E(3)-E(5)+E(7)-\ldots\right]. \end{aligned} \tag{144}$$

Now, by choosing different correction-divisors a_p in (144), we get several transformed series which have better convergence properties. If we consider the correction-divisor (136), then using the expression (138) for the *sthaulyas*, we get

$$\begin{aligned} C &= 4d\left(1-\frac{1}{5}\right)-16d\left[\frac{1}{(3^5+4.3)}-\frac{1}{(5^5+4.5)}+\frac{1}{(7^5+4.7)}-\ldots\right] \\ &= 16d\left[\frac{1}{(1^5+4.1)}-\frac{1}{(3^5+4.3)}+\frac{1}{(5^5+4.5)}-\ldots\right]. \end{aligned} \tag{145}$$

The above series is given in the verse *samapañcāhatayoḥ* ...(I). Note that each term in the above series involves the fifth power of the odd number in the denominator, unlike the original series which only involved the first power of the odd number. Clearly, this transformed series gives more accurate results with fewer terms.

If we had used only the lowest order correction (132) and the associated *sthaulya* (134), instead of the correction employed above, then the transformed series is the one given in the verse *vyāsād vāridhinihatāt*...(II)

$$C = 4d\left[\frac{3}{4}+\frac{1}{(3^3-3)}-\frac{1}{(5^3-5)}+\frac{1}{(7^3-7)}-\ldots\right]. \tag{146}$$

Note that the denominators in the above transformed series are proportional to the third power of the odd number.

Even if we take non-optimal correction-divisors, we often end-up obtaining interesting series. For instance, if we take a non-optimal correction-divisor, say of the form $a_p = 2p$, then the *sthaulya* is given by

$$\begin{aligned} E(p) &= \frac{1}{(2p-4)} + \frac{1}{2p} - \frac{1}{p} \\ &= \frac{1}{(p^2-2p)} \\ &= \frac{1}{(p-1)^2-1}. \end{aligned} \tag{147}$$

Then, the transformed series will be the one given in the verse *dvyādiyujāṃ vā kṛtayo*. . .(III)[92]

$$C = 4d\left[\frac{1}{2} + \frac{1}{(2^2-1)} - \frac{1}{(4^2-1)} + \frac{1}{(6^2-1)} + \ldots\right]. \tag{148}$$

16. Derivation of the Mādhava series for Rsine and Rversine

16.1. First and second order differences of Rsines

We shall now outline the derivation of Mādhava series for Rsine (*bhujā-jyā*) and Rversine (*śara*), as given in *Yuktibhāṣā*.[93] *Yuktibhāṣā* begins with a discussion of the first and second order Rsine-differences and derives an exact form of the result of Āryabhaṭa that the second-order Rsine-differences are proportional to the Rsines themselves. We had briefly indicated this proof in Section 5.3.

Here we are interested in obtaining the Mādhava series for the *jyā* and *śara* of an arc of length s indicated by EC in Figure 9. This arc is divided into n equal arc bits, where n is large. If the arc length $s = R\theta$, then the j-th *piṇḍa-jyā*, B_j is given by[94]

$$B_j = jy\bar{a}\left(\frac{js}{n}\right) = R\sin\left(\frac{j\theta}{n}\right). \tag{149}$$

---

[92]The verse III in fact presents the series (148) along with an end correction-term of the form $(-1)^p \frac{4d}{2(p+1)^2+2}$.

[93]*Yuktibhāṣā*, cited earlier, Vol. I Section 16.5, pp. 94–103, 221–233, 417–427.

[94]Figure 9 is essentially the same as Figure 3 considered in section 5 except that the *piṇḍajyās* B_j are Rsines associated with multiples of the arc-bit $\frac{s}{n}$ into which the arc $EC = s$ is divided. In Figure 3, the B_j's are the tabular Rsines associated with multiples of 225′.

The corresponding *koṭi-jyā* K_j, and the *śara* S_j, are given by

$$K_j = koṭi\left(\frac{js}{n}\right) = R\cos\left(\frac{j\theta}{n}\right), \tag{150}$$

$$S_j = śara\left(\frac{js}{n}\right) = R\left[1-\cos\left(\frac{j\theta}{n}\right)\right]. \tag{151}$$

Now, C_jC_{j+1} represents the $(j+1)$-th arc bit. Then, for the arc $EC_j = \frac{js}{n}$, its *piṇḍa-jyā* is $B_j = C_jP_j$, and the corresponding *koṭi-jyā* and *śara* are $K_j = C_jT_j$, $S_j = EP_j$. Similarly we have

$$B_{j+1} = C_{j+1}P_{j+1},\ K_{j+1} = C_{j+1}T_{j+1} \text{ and } S_{j+1} = EP_{j+1}. \tag{152}$$

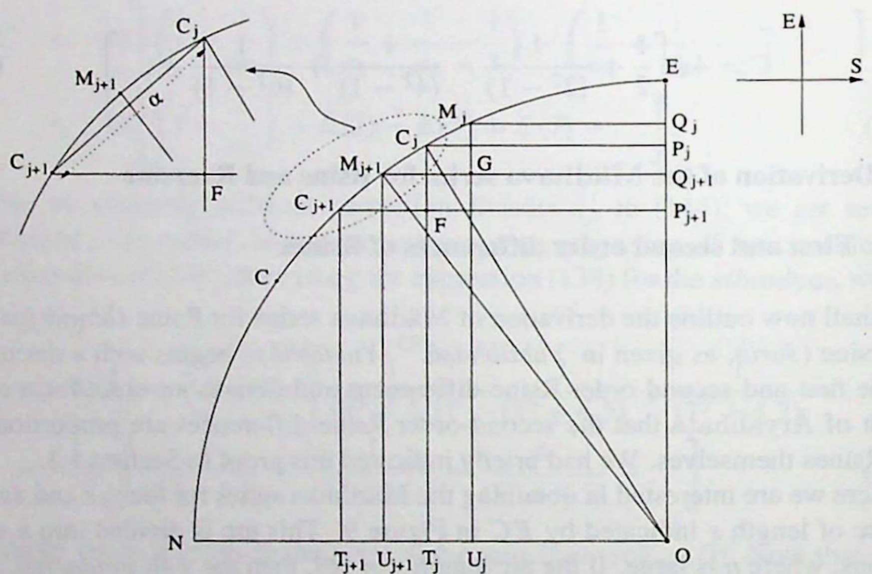


FIGURE 9. Computation of *Jyā* and *Śara* by *Saṅkalitas.*

Let M_{j+1} be the mid-point of the arc-bit C_jC_{j+1} and similarly M_j the mid-point of the previous (j-th) arc-bit. We shall denote the *piṇḍa-jyā* of the arc EM_{j+1} as $B_{j+\frac{1}{2}}$ and clearly

$$B_{j+\frac{1}{2}} = M_{j+1}Q_{j+1}\ .$$

The corresponding *koṭi-jyā* and *śara* are

$$K_{j+\frac{1}{2}} = M_{j+1}U_{j+1} \text{ and } S_{j+\frac{1}{2}} = EQ_{j+1}\ .$$

Similarly,

$$B_{j-\frac{1}{2}} = M_j Q_j, \; K_{j-\frac{1}{2}} = M_j U_j \text{ and } S_{j-\frac{1}{2}} = E Q_j \,. \tag{153}$$

Let α be the chord corresponding to the equal arc-bits $\frac{s}{n}$ as indicated in Figure 9. That is, $C_j C_{j+1} = M_j M_{j+1} = \alpha$. Let F be the intersection of $C_j T_j$ and $C_{j+1} P_{j+1}$, and G of $M_j U_j$ and $M_{j+1} Q_{j+1}$. The triangles $C_{j+1} F C_j$ and $O Q_{j+1} M_{j+1}$ are similar, as their sides are mutually perpendicular. Thus we have

$$\frac{C_{j+1} C_j}{O M_{j+1}} = \frac{C_{j+1} F}{O Q_{j+1}} = \frac{F C_j}{Q_{j+1} M_{j+1}} \,. \tag{154}$$

Hence we obtain

$$B_{j+1} - B_j \;=\; \left(\frac{\alpha}{R}\right) K_{j+\frac{1}{2}}, \tag{155}$$

$$K_j - K_{j+1} \;=\; S_{j+1} - S_j = \left(\frac{\alpha}{R}\right) B_{j+\frac{1}{2}}. \tag{156}$$

Similarly, the triangles $M_{j+1} G M_j$ and $O P_j C_j$ are similar and we get

$$\frac{M_{j+1} M_j}{O C_j} = \frac{M_{j+1} G}{O P_j} = \frac{G M_j}{P_j C_j}. \tag{157}$$

Thus we obtain

$$B_{j+\frac{1}{2}} \;-\; B_{j-\frac{1}{2}} = \left(\frac{\alpha}{R}\right) K_j, \tag{158}$$

$$K_{j-\frac{1}{2}} \;-\; K_{j+\frac{1}{2}} = S_{j+\frac{1}{2}} - S_{j-\frac{1}{2}} = \left(\frac{\alpha}{R}\right) B_j \,. \tag{159}$$

We define the Rsine-differences (*khaṇḍa-jyā*) Δ_j by

$$\Delta_j = B_j - B_{j-1} \,, \tag{160}$$

with the convention that $\Delta_1 = B_1$. From (155), we have

$$\Delta_j = \left(\frac{\alpha}{R}\right) K_{j-\frac{1}{2}} \,. \tag{161}$$

From (159) and (161), we also get the second order Rsine-differences (the differences of the Rsine-differences called *khaṇḍa-jyāntara*):

$$\begin{aligned}\Delta_j - \Delta_{j+1} &= (B_j - B_{j-1}) - (B_{j+1} - B_j) \\ &= \left(\frac{\alpha}{R}\right)\left(K_{j-\frac{1}{2}} - K_{j+\frac{1}{2}}\right) \\ &= \left(\frac{\alpha}{R}\right)\left(S_{j+\frac{1}{2}} - S_{j-\frac{1}{2}}\right) \\ &= \left(\frac{\alpha}{R}\right)^2 B_j \,. \end{aligned} \qquad (162)$$

Now, if the sum of the second-order Rsine-differences, is subtracted from the first Rsine-difference, then we get any desired Rsine-difference. That is

$$\Delta_1 - \left[(\Delta_1 - \Delta_2) + (\Delta_2 - \Delta_3) + \ldots + (\Delta_{j-1} - \Delta_j)\right] = \Delta_j \,. \qquad (163)$$

From (162) and (163) we conclude that

$$\Delta_1 - \left(\frac{\alpha}{R}\right)^2 (B_1 + B_2 + \cdots + B_{j-1}) = \Delta_j \,. \qquad (164)$$

16.2. Rsines and Rversines from *Jyā-saṅkalita*

We can sum up the Rversine-differences (159), to obtain the *śara*, Rversine, at the midpoint of the last arc-bit as follows:

$$\begin{aligned} S_{n-\frac{1}{2}} - S_{\frac{1}{2}} &= \left(S_{n-\frac{1}{2}} - S_{n-\frac{3}{2}}\right) + \ldots\ldots \left(S_{\frac{3}{2}} - S_{\frac{1}{2}}\right) \\ &= \left(\frac{\alpha}{R}\right)(B_{n-1} + B_{n-2} + \ldots + B_1) \,. \end{aligned} \qquad (165)$$

Using (162), the right hand side of (165) can also be expressed as a summation of the second order differences. From (164) and (165) it follows that the Rversine at the midpoint of the last arc-bit is also given by

$$\left(\frac{\alpha}{R}\right)\left(S_{n-\frac{1}{2}} - S_{\frac{1}{2}}\right) = (\Delta_1 - \Delta_n). \qquad (166)$$

Now, since the first Rsine-difference $\Delta_1 = B_1$, any desired Rsine can be obtained by adding the Rsine-differences; these Rsine-differences have been obtained in (164). Now, by making use of (164), the last *piṇḍa-jyā* can be expressed as follows:

$$B_n = \Delta_n + \Delta_{n-1} + \ldots + \Delta_1$$

$$= n\Delta_1 - \left(\frac{\alpha}{R}\right)^2 \left[(B_1 + B_2 \ldots + B_{n-1}) + (B_1 + B_2 \ldots + B_{n-2}) + \ldots + B_1\right]$$

$$= nB_1 - \left(\frac{\alpha}{R}\right)^2 \left[B_{n-1} + 2B_{n-2} + \cdots + (n-1)B_1\right]. \tag{167}$$

The results (158) – (167), obtained so far, involve no approximations. It is now shown how better and better approximations to the Rsine and Rversine can be obtained by taking n to be very large or, equivalently, the arc-bit $\frac{s}{n}$ to be very small. Then, we can approximate the full-chord and the Rsine of the arc-bit by the arc-bit $\frac{s}{n}$ itself. Also, as a first approximation, we can approximate the *piṇḍa-jyās* B_j in the equations (164), (165) or (167) by the corresponding arcs themselves. That is

$$B_j \approx \frac{js}{n}. \tag{168}$$

The result for the Rsine obtained this way is again used to obtain a better approximation for the *piṇḍa-jyās* B_j which is again substituted back into the equations (165) and (167) and thus by a process of iteration successive better approximations are obtained for the Rsine and Rversine. Now, once we take $B_j \approx \frac{js}{n}$, we will be led to estimate the sums and repeated sums of natural numbers (*ekādyekottara-saṅkalita*), when the number of terms is very large.

16.3. Derivation of Mādhava series by iterative corrections to *jyā* and *śara*

As we noted earlier, the relations given by (165) and (167) are exact. But now we shall show how better and better approximations to the Rsine and Rversine of any desired arc can be obtained by taking n to be very large or, equivalently, taking the arc-bit $\frac{s}{n}$ to be very small. Then both the full-chord α, and the first Rsine B_1 (the Rsine of the arc-bit), can be approximated by the arc-bit $\frac{s}{n}$ itself, and the Rversine $S_{n-\frac{1}{2}}$ can be taken as S_n and the Rversine $S_{\frac{1}{2}}$ may be treated as negligible. Thus

the above relations (165), (167) become[95]

$$S = S_n \approx \left(\frac{s}{nR}\right)(B_{n-1} + B_{n-2} + \ldots + B_1), \tag{169}$$

$$B = B_n \approx s - \left(\frac{s}{nR}\right)^2 [(B_1 + B_2 + \ldots + B_{n-1}) + (B_1 + B_2 \ldots + B_{n-2}) + \ldots + B_1], \tag{170}$$

where B and S are the Rsine and Rversine of the desired arc of length s and the results will be more accurate, larger the value of n.

Now, as a first approximation, we take each *piṇḍa-jyā* B_j in (169) and (170) to be equal to the corresponding arc itself, that is

$$B_j \approx \frac{js}{n}. \tag{171}$$

Then we obtain for the Rversine

$$S \approx \left(\frac{s}{nR}\right)\left[(n-1)\left(\frac{s}{n}\right) + (n-2)\left(\frac{s}{n}\right) + \ldots\right] = \left(\frac{1}{R}\right)\left(\frac{s}{n}\right)^2 [(n-1) + (n-2) + \ldots]. \tag{172}$$

For large n, we can use the estimate (89) for the sum of integers. Hence (172) reduces to

$$S \approx \left(\frac{1}{R}\right)\frac{s^2}{2}. \tag{173}$$

Equation (173) is the first *śara-saṃskāra*, correction to the Rversine. We now substitute our first approximation (171) to the *piṇḍa-jyās* B_j in (170), which gives the Rsine of the desired are as a second order repeated sum of the *piṇḍa-jyās* B_j. We then obtain

$$B \approx s - \left(\frac{1}{R}\right)^2\left(\frac{s}{n}\right)^3 [(1+2+\ldots+(n-1)) + (1+2+\ldots(n-2)) + \ldots]. \tag{174}$$

[95] As has been pointed out by one of the reviewers, in the following derivation instead of using the relation (170), which involves repeated summation of *piṇḍajyās*, one could use the much simpler relation

$$B = B_n \approx s - \frac{s}{nR}(S_{n-1} + S_{n-2} + \ldots + S_1),$$

which essentially follows from (165) and (170). Then we can iterate between the above equation and (169) which involve considering only sums of powers of integers. *Yuktibhāṣā*, however, employes successive iteration between (169) and (170), which involves consideration of repeated sums of integers.

The second term in (174) is a *dvitīya-saṅkalita*, the second order repeated sum, and using the estimate (108), we obtain

$$B \approx s - \left(\frac{1}{R}\right)^2 \frac{s^3}{1.2.3}\,. \tag{175}$$

Thus we see that the first correction obtained in (175) to the Rsine-arc-difference (*jyā-cāpāntara-saṃskāra*), is equal to the earlier correction to the Rversine (*śara-saṃskāra*) given in (173) multiplied by the arc and divided by the radius and 3.

It is noted that the results (173) and (175) are only approximate (*prāyika*), since, instead of the *saṅkalita* of the *piṇḍa-jyās* in (169) and (170), we have only carried out *saṅkalita* of the arc-bits. Now that (175) gives a correction to the difference between the Rsine and the arc (*jyā-cāpāntara-saṃskāra*), we can use that to correct the values of the *piṇḍa-jyās* and thus obtain the next corrections to the Rversine and Rsine.

Following (175), the *piṇḍa-jyās* may now be taken as

$$B_j \approx \frac{js}{n} - \left(\frac{1}{R}\right)^2 \left[\frac{\left(\frac{js}{n}\right)^3}{1.2.3}\right]. \tag{176}$$

If we introduce (176) in (169), we obtain

$$\begin{aligned} S \;\approx\; & \left(\frac{1}{R}\right)\left(\frac{s}{n}\right)^2 [(n-1)+(n-2)+\ldots] \\ & - \left(\frac{s}{nR}\right)\left(\frac{1}{R}\right)^2\left(\frac{s}{n}\right)^3\left(\frac{1}{1.2.3}\right)[(n-1)^3+(n-2)^3+\ldots]. \end{aligned} \tag{177}$$

The first term in (177) was already evaluated while deriving (173). The second term in (177) can either be estimated as a summation of cubes (*ghana-saṅkalita*), or as a *tṛtīya-saṅkalita*, third order (repeated) summation, because each individual term there has been obtained by doing a second-order (repeated) summation. Hence, recollecting our earlier estimate (110) for these *saṅkalitas*, we get

$$S \approx \left(\frac{1}{R}\right)\frac{s^2}{1.2} - \left(\frac{1}{R}\right)^3 \frac{s^4}{1.2.3.4}\,. \tag{178}$$

Equation (178) gives a correction (*śara-saṃskāra*) to the earlier value (173) of the Rversine; which is nothing but the earlier correction to the Rsine-arc difference (*jyā-cāpāntara-saṃskāra*) given in (175) multiplied by the arc and divided by the radius and 4.

Again, if we use the corrected *piṇḍa-jyās* (176) in the expression (170) for the Rsine, we obtain

$$\begin{aligned} B \approx\ & s - \left(\frac{1}{R}\right)^2 \left(\frac{s}{n}\right)^3 [(1+2+..+(n-1)) + (1+2+..+(n-2)) + ..] \\ & + \left(\frac{1}{R}\right)^4 \left(\frac{s}{n}\right)^5 \\ & \times \left(\frac{1}{1.2.3}\right) \left[(1^3+2^3+...+(n-1)^3) + (1^3+2^3+...+(n-2)^3) + ..\right] \\ \approx\ & s - \left(\frac{1}{R}\right)^2 \frac{s^3}{1.2.3} + \left(\frac{1}{R}\right)^4 \frac{s^5}{1.2.3.4.5}\,. \end{aligned} \tag{179}$$

The above process can be repeated to obtain successive higher order corrections for the Rversine and Rsine: By first finding a correction (*jyā-cāpāntara-saṃskāra*) for the difference between the Rsine and the arc, using this correction to correct the *piṇḍa-jyās* B_j, and using them in equations (169) and (170) get the next correction (*śara-saṃskāra*) for the Rversines, and the next correction (*jyā-cāpāntara-saṃskāra*) for the Rsine-arc-difference itself, which is then employed to get further corrections iteratively. In this way we are led to the Mādhava series for *jyā* and *śara* given by

$$\begin{aligned} B = R\sin(s) &= s - \left(\frac{1}{R}\right)^2 \frac{s^3}{(1.2.3)} + \left(\frac{1}{R}\right)^4 \frac{s^5}{(1.2.3.4.5)} \\ &\quad - \left(\frac{1}{R}\right)^6 \frac{s^7}{(1.2.3.4.5.7)} + \dots, \\ S = R\ \text{vers}(s) &= \left(\frac{1}{R}\right) \frac{s^2}{2} - \left(\frac{1}{R}\right)^3 \frac{s^4}{(1.2.3.4)} \\ &\quad + \left(\frac{1}{R}\right)^5 \frac{s^6}{(1.2.3.4.6)} - \dots \,. \end{aligned} \tag{180}$$

That is,

$$\begin{aligned} \sin\theta &= \theta - \frac{\theta^3}{(1.2.3)} + \frac{\theta^5}{(1.2.3.4.5)} - \frac{\theta^7}{(1.2.3.4.5.6.7)} + \dots\,, \\ \text{vers}\,\theta &= \frac{\theta^2}{(1.2)} - \frac{\theta^4}{(1.2.3.4)} + \frac{\theta^6}{(1.2..4.5.6)} - \dots\,. \end{aligned} \tag{181}$$

17. Instantaneous velocity and derivatives

As we saw in Section 6.1, the *mandaphala* or the equation of centre for a planet $\Delta\mu$ is given by

$$R\sin(\Delta\mu) = \left(\frac{r_0}{R}\right) R\sin(M-\alpha), \tag{182}$$

where r_0 is the mean epicycle radius, M is the mean longitude of the planet and α the longitude of the apogee. Further as we noted earlier, Muñjāla, Āryabhaṭa II and Bhāskara II used the approximation

$$R\sin(\Delta\mu) \approx \Delta\mu, \tag{183}$$

in (182) and obtained the following expression as correction to the instantaneous velocity of the planet:

$$\frac{d}{dt}(\Delta\mu) = \left(\frac{r_0}{R}\right) R\cos(M-\alpha)\frac{d}{dt}(M-\alpha). \tag{184}$$

Actually the instantaneous velocity of the planet has to be evaluated from the more accurate relation

$$\Delta\mu = R\sin^{-1}\left[\left(\frac{r_0}{R}\right) R\sin(M-\alpha)\right]. \tag{185}$$

The correct expression for the instantaneous velocity which involves the derivative of arc-sine function has been given by Nīlakaṇṭha in his *Tantrasaṅgraha.*[96]

चन्द्रबाहुफलवर्गशोधितत्रिज्यकाकृतिपदेन संहरेत्।
तत्र कोटिफललिप्तिकाहतां केन्द्रभुक्तिरिह यच्च लभ्यते॥
तद्विशोध्य मृगादिके गतेः क्षिप्यतामिह तु कर्कटादिके।
तद्भवेत्स्फुटतरा गतिर्विधोः अस्य तत्समयजा रवेरपि॥

Let the product of the *koṭiphala* [$r_0\cos(M-\alpha)$] in minutes and the daily motion of the *manda-kendra* $\left(\frac{d(M-\alpha)}{dt}\right)$ be divided by the square root of the square of the *bāhuphala* subtracted from the square of *trijyā* $\left(\sqrt{R^2 - r_0^2\sin^2(M-\alpha)}\right)$. The result thus obtained has to be subtracted from the daily motion of the Moon if the *manda-kendra* lies within six signs beginning from *Mṛga* and added if it

[96] *Tantrasaṅgraha*, cited above (fn. 52), verses 2.53–54, pp.169–170. Elsewhere, Nīlakaṇṭha has ascribed these verses to his teacher Dāmodara (*Jyotirmīmāṃsā*, Ed. by K. V. Sarma, VVRI, Hoshiarpur 1977, p. 40).

lies within six signs beginning from *Karkaṭaka*. The result gives a more accurate value of the Moon's angular velocity. In fact, the procedure for finding the instantaneous velocity of the Sun is also the same.

If $(M - \alpha)$ be the *manda-kendra*, then the content of the above verse can be expressed as

$$\frac{d}{dt}\left[\sin^{-1}\left(\frac{r_0}{R}\sin(M-\alpha)\right)\right] = \frac{r_0\cos(M-\alpha)\,\dfrac{d(M-\alpha)}{dt}}{\sqrt{R^2 - r_0^2\sin^2(M-\alpha)}}. \quad (186)$$

The instantaneous velocity of the planet is given by

$$\frac{d}{dt}\mu = \frac{d}{dt}(M-\alpha) - \frac{r_0\cos(M-\alpha)\,\dfrac{d(M-\alpha)}{dt}}{\sqrt{R^2 - r_0^2\sin^2(M-\alpha)}}. \quad (187)$$

Here, the first term in the RHS represents the mean velocity of the planet and the second term the rate of change in the *mandaphala* given by (186).

In his *Āryabhaṭīya-bhāṣya*, Nīlakaṇṭha explains how his result is more correct than the traditional result of Muñjāla and Bhāskarācārya:[97]

अतः फलसाम्यं कुतः? ...पुनरपि यो विशेषः तत्र कोटिज्यागुणितस्य त्रिज्यया हरणमुक्तम्, इह कोटिफलगुणितस्य केन्द्रभोगस्य दोःफलकोट्या हरणमुक्तम् इति। तेन तत्फलं चापीकृतं भुजाफलगतिः स्यात्। कथम्?

चापगतिसम्बन्धिज्यागत्यानयने यत् त्रैराशिकमुक्तं, ज्यागत्या चापगत्यानयने तद्विपरीतं कर्म कार्यम्। तत्र पूर्वोक्ते कर्मणि त्रैराशिकद्वयेन या दोःफलगतिः आनीता तां व्यासार्धेन हत्वा दोःफलकोट्या हृत्वा तच्चापगतिर्लभ्या। तत्रेदं त्रैराशिकम् ...

Hence, how can the results be equal? ... Again the distinction being: there it was prescribed that the multiplier *koṭi-jyā* was to be divided by *trijyā*, [but] here it has been prescribed that the product of *koṭiphala* and the rate of change of *kendra* be divided by *koṭi* of the *doḥphala* (*doḥphalakoṭyā*).[98] ...

---

[97] *Āryabhaṭīya* of Āryabhaṭa, Ed. with *Bhāṣya* of Nīlakaṇṭha Somayājī by K. Sāmbaśiva Śāstrī, Trivandrum Sanskrit Series 110, Trivandrum 1931, comm. on *Kālakriyāpāda* 22–25, pp. 62–63.

[98] The terms *doḥphala* and *koṭiphala* refer to $\frac{r_0}{R}\sin(M-\alpha)$ and $\frac{r_0}{R}\cos(M-\alpha)$ respectively. Hence, the term *doḥphalakoṭi* refers to $\sqrt{1-(\frac{r_0}{R}\sin(M-\alpha))^2}$.

17.1. Acyuta's expression for instantaneous velocity involving the derivative of ratio of two functions

In the third chapter of his *Sphuṭanirṇayatantra*, Acyuta Piṣāraṭi (c. 1550–1621), a disciple of Jyeṣṭhadeva, discusses various results for the instantaneous velocity of a planet depending on the form of equation of centre (*manda-saṃskāra*). He first presents the formula involving the derivative of arc-sine function given by Nīlakaṇṭha (in the name of (*manda*)-*sphuṭagati*) as follows:[99]

कोटिफलाहतकेन्द्रगतेर्यद् दोःफलकोटिकयाप्तमनेन।
हीनयुतामृगकर्कटकाद्योर्मध्यगतिर्भवति स्फुटभुक्तिः ॥

Acyuta also gives the formula for the instantaneous velocity of a planet if one were to follow a different model proposed by Munjāla for the equation of centre, according to which *mandaphala* is given by

$$\Delta\mu = \frac{\frac{r_0}{R}\sin(M-\alpha)}{\left(1-\frac{r_0}{R}\cos(M-\alpha)\right)}, \tag{188}$$

instead of (182), where $\Delta\mu$ is small. If one were to use this formula for *mandaphala* for finding the true longitude of the planet, then it may be noted that the instantaneous velocity will involve the derivative of the ratio of two functions both varying with time. Taking note of this, Acyuta observes:[100]

कृत्स्नस्य मान्दपरिधेर्निजकर्णतुल्यौ
वृद्धिक्षयाविति मते कथितः क्रमोऽयम्।
अर्धस्य मान्दपरिधेः क्षयवृद्धिपक्षे
युक्तं क्रियाक्रममथ प्रतिपादयामः ॥

The procedure that was prescribed earlier is with reference to the School that conceives of the increase and decrease in the circumference of the *manda-vṛtta* in accordance with the *karṇa*. With reference to the School that conceives of increase and decrease only according to the half [of it], now we prescribe the appropriate procedure to be adopted.

Acyuta then proceeds to give the correct expression for the instantaneous velocity of a planet in Munjāla's model:[101]

---

[99] *Sphuṭanirṇayatantra* of Acyuta Piṣāraṭi, Ed. by K. V. Sarma, VVRI, Hoshiarpur 1974, p. 19.

[100] Ibid., p. 20.

[101] Ibid., p. 21.

कृतकोटिफलं त्रिजीवया विहृतं दोःफलवर्गतस्तु यत्।
मृगकर्कटकादिकेऽमुना युतहीनं फलमत्रकोटिजम्॥
दिनकेन्द्रगतिघ्नमुद्धरेत् कृतकोटीफलया त्रिजीवया।
फलपूर्वफलैकतो दलं दिनभुक्तेरपि संस्कृतिर्भवेत्॥

Having applied the *koṭiphala* to *trijyā* [positively or negatively depending upon the *mandakendra*], let the square of the *doḥphala* be divided by that. This may be added to or subtracted from the *koṭiphala* depending on whether it is *Mṛgādi* or *Karkyādi*. The product of this [result thus obtained] and the daily motion of the *manda-kendra* divided by the *koṭiphala* and applied to *trijyā* will be the correction to the daily motion.

Thus according to Acyuta, the correction to the mean velocity of a planet in order to obtain its instantaneous velocity is given by

$$\frac{\left(\frac{r_0}{R}\cos(M-\alpha)\right)+\frac{\left(\frac{r_0}{R}\sin(M-\alpha)\right)^2}{\left(1-\frac{r_0}{R}\cos(M-\alpha)\right)}}{\left(1-\frac{r_0}{R}\cos(M-\alpha)\right)}\,\frac{d(M-\alpha)}{dt}, \qquad (189)$$

which is nothing but the derivative of the expression given in (188).

Acknowledgment

We are thankful to Prof. D. Mumford and Prof. C. S. Seshadri for organizing the seminar on History of Mathematics at Chennai Mathematical Institute during January–February 2008, which gave us an opportunity to put together some of the material on development of calculus in India. We are grateful to our collaborator Prof. M. S. Sriram and also to Prof. R. Sridharan and Prof. Raja Sridharan for their valuable suggestions and continued encouragement. We would like to specially thank the reviewers of the paper for their valuable comments and suggestions.

References

Major Source Works

(1) *The Śulbasūtras of Baudhāyana, Āpastamba, Kātyāyana and Mānava*: Ed. and Tr. by S. N. Sen and A. K. Bag, Indian National Science Academy, New Delhi 1983.

(2) *Āryabhaṭīya* of Āryabhaṭa (c. 499): Ed. and Tr. by K. S. Shukla and K. V. Sarma, Indian National Science Academy, New Delhi 1976.

(3) *Āryabhaṭīyabhāṣya* (c. 629) of Bhāskara I: Ed. by K. S. Shukla, Indian National Science Academy, New Delhi 1976. Eng Tr. of *Gaṇitapāda* by Agathe Keller, 2 Vols. Birkhauser 2006.

(4) *Brāhmasphuṭasiddhānta* (c. 628) of Brahmagupta: Mathematics Chapters (XIII, XVIII) translated in H. T. Colebrooke, *Algebra with Arithmetic and Mensuration from the Sanskrit of Brahmagupta and Bhāskara*, London 1817. Ed. by Sudhakara Dvivedi, Varanasi 1902.

(5) *Bakṣāli Manuscript* (c. 8th cent): Ed. and Tr. by T. Hayashi, in *The Bakṣāli Manuscript: An Ancient Indian Mathematical Treatise*, Egbert Forsten, Groningen 2005.

(6) *Triśatikā* (c. 750) of Śrīdhara: Ed. by Sudhakara Dvivedi, Benaras 1899. Tr. by N. Ramanujachari and G. R. Kaye, Bibliotheca Mathematica, **13**, 203–217, 1912.

(7) *Bhāṣya* (c. 800) of Govindasvāmin on *Mahābhāskarīya* of Bhāskara I: Ed. by T. S. Kuppanna Sastri, Madras 1957.

(8) *Gaṇitasārasaṅgraha* (c. 850) of Mahāvīra: Ed. and Tr. by M. Rangacharya, Govt. Press, Madras 1912. Ed. with Hindi Tr. by L C Jain, Jain Samskrit Samrakshaka Sangha, Sholapur 1963.

(9) *Vāsanā* (c. 860) of Pṛthūdakasvāmin on *Brāhmasphuṭasiddhānta* of Brahmagupta: partially Ed. by Ramaswarup Sharma, 4 Vols, New Delhi 1966; Chapter 21, Ed. and Tr. by S. Ikeyama, INSA, New Delhi 2003.

(10) *Līlāvatī* (c. 1150) of Bhāskarācārya II: Tr. in H. T. Colebrooke, *Algebra with Arithmetic and Mensuration from the Sanskrit of Brahmagupta and Bhāskara*, London 1817. Ed. with Bhāskara's *Vāsanā* and *Buddhivilāsinī* of Gaṇeśa Daivajña (c. 1545) by V. G. Apte, 2 Vols, Pune 1937.

(11) *Bījagaṇita* (c. 1150) of Bhāsakarācārya II: Tr. in H. T. Colebrooke, *Algebra with Arithmetic and Mensuration from the Sanskrit of Brahmagupta and Bhāskara*, London 1817. Ed. with Bhāskara's *Vāsanā* by Muralidhara Jha, Benares 1927, Chaukhamba Rep. 1994.

(12) *Siddhāntaśiromaṇi* (c. 1150) of Bhāskarācārya II: Ed. with Bhāskara's *Vāsanā* by Bapudeva Sastri, Varanasi 1866. Ed. with Bhāskara's *Vāsanā* and Nṛsiṃha Daivajña's *Vāsanāvārttika* by Muralidhara Chaturveda, Varanasi 1981. *Grahagaṇitādhyāya* Tr. by D. Arkasomayaji, Kendriya Sanskrit Vidya-peetha, Tirupati 1980, Rep. 2000.

(13) *Gaṇitakaumudī* (c. 1356) of Nārāyaṇa Paṇḍita: Ed. by Padmakara Dvivedi, 2 Vols, Varanasi 1936, 1942. Chapters 13–14 Ed. and Tr. by T Kusuba, *Combinatorics and Magic-squares in India: A Study of Nārāyaṇa Paṇḍita's Gaṇitakaumudī, Chapters 13–14*, PhD Dissertation, Brown University 1993. Tr. by Parmanand Singh, *Gaṇitabhāratī* 1998–2002.

(14) *Tantrasaṅgraha* (c. 1500) of Nīlakaṇṭha Somayājī: Ed. with *Laghuvivṛti* of Śaṅkara Vāriyar by Surnad Kunjan Pillai, Trivandrum 1958. Ed. with *Yuktidīpikā* of Śaṅkara Vāriyar by K. V. Sarma, Hoshiarpur 1977.

(15) *Āryabhaṭīyabhāṣya* (c. 1502) of Nīlakaṇṭha Somayājī on *Āryabhaṭīya* of Āryabhaṭa: Vols. 1 and 2 on *Gaṇita* and *Kālakriyāpāda*, Ed. by K. *Sāmbaśiva Śastrī*, Trivandrum 1930, 1931; Vol. 3 on *Golapāda*. Ed. by Surnad Kunjan Pillai, Trivandrum 1957.

(16) *Gaṇita-yukti-bhāṣā* (c. 1530) of Jyeṣṭhadeva (in Malayalam): *Gaṇitādhyāya*, Ed., with Notes in Malayalam, by Ramavarma Thampuran and A. R. Akhileswarayyar, Trichur 1948. Ed. with Tr. by K. V. Sarma with Explanatory Notes by K. Ramasubramanian, M. D. Srinivas and M. S. Sriram, 2 Vols, Hindustan Book Agency, Delhi 2008, Rep. Springer 2009.

(17) *Kriyākramakarī* (c. 1535) of Śaṅkara Vāriyar on *Līlāvatī* of Bhāskarācārya II: Ed. by K. V. Sarma, Hoshiarpur 1975.

(18) *Yuktidīpikā* (c. 1535) of Śaṅkara Vāriyar on *Tantrasaṅgraha* of Nīlakaṇṭha Somayājī: Ed. by K. V. Sarma, Hoshiarpur 1977.

(19) *Buddhivilāsinī* (c. 1545) of Gaṇeśa Daivajña on *Līlāvatī* of Bhāskarācārya II: Ed. by V. G. Apte, 2 Vols, Pune 1937.

(20) *Bījanavāṅkura* (c. 1600) of Kṛṣṇa Daivajña on *Bījagaṇita* of Bhāskarācār-ya II: Ed. by V. G. Apte, Pune 1930. Ed. by T. V. Radha Krishna Sastry, Tanjore 1958.

(21) *Sphuṭanirṇayatantra* (c. 1600) of Acyuta Piśāraṭi: Ed. with Acyuta's commentary by K. V. Sarma, Hoshiarpur 1974.

Some Overviews of Indian Mathematics

(1) B. B. Datta and A. N. Singh, *History of Hindu Mathematics*, 2 parts, Lahore 1935, 1938. Rep. Bharatiya Kala Prakasan, Delhi 2004.

(2) C. N. Srinivasa Iyengar, *History of Indian Mathematics*, Calcutta 1967.

(3) A. K. Bag, *Mathematics in Ancient and Medieval India*, Varanasi 1979.

(4) T. A. Sarasvati Amma, *Geometry in Ancient and Medieval India*, Delhi 1979. Rep. 2007.

(5) G. G. Joseph, *The Crest of the Peacock: The Non-European Roots of Mathematics*, Peguin 1990, 2nd Ed., Princeton 2000.

(6) S. Parameswaran, *The Golden Age of Indian Mathematics*, Swadeshi Science Movement, Kochi 1998.

(7) G. Emch, M. D. Srinivas and R. Sridharan Eds., *Contributions to the History of Mathematics in India*, Hindustan Book Agency, Delhi 2005.

(8) C. K. Raju, *Cultural Foundations of Mathematics: The Nature of Mathematical Proof and the Transmission of the Calculus from India to Europe in the* 16^{th}*c.CE*, Pearson Education, Delhi 2007.

(9) Kim Plofker, 'Mathematics in India', in V. J. Katz Ed., *The Mathematics of Egypt, Mesopotamia, China, India and Islam: A Source Book*, Princeton 2007, pp. 385–513.

(10) Kim Plofker, *History of Mathematics in India: From 500 BCE to 1800 CE*, Princeton 2008.

Some Articles

(1) C. M. Whish, 'On the Hindu quadrature of the circle, and the infinite series of the proportion of the circumference to the diameter exhibited in the four Shastras, the Tantrasangraham, Yucti Bhasa, Carana Paddhati and Sadratnamala', Tran. Roy. As. Soc. (G.B.) **3**, 509–523, 1834.

(2) Bapudeva Sastri, 'Bhaskara's knowledge of the differential calculus', Jour. As. Soc. Bengal **27**, 213–216, 1858.

(3) William Spotiswoode, 'Note on the supposed discovery of the principle of the differential calculus by an Indian astronomer', Jour. Royal As. Soc. **17**, 221–222, 1860.

(4) P. C. Sen Gupta, 'Infinitesimal calculus in Indian mathematics', Jour. Dep. Letters. Calcutta Univ., **22**, 1–17, 1931.

(5) K. Mukunda Marar, 'Proof of Gregory's series', Teacher's Magazine **15**, 28–34, 1940.

(6) K. Mukunda Marar and C. T. Rajagopal, 'On the Hindu quadrature of the circle', J.B.B.R.A.S. **20**, 65–82, 1944.

(7) K. Mukunda Marar and C. T. Rajagopal, 'Gregory's series in the mathematical literature of Kerala', Math. Student **13**, 92–98, 1945.

(8) A. Venkataraman, 'Some interesting proofs from *Yuktibhāṣā*', Math. Student **16**, 1–7, 1948.

(9) C. T. Rajagopal, 'A neglected chapter of Hindu mathematics', Scr. Math. **15**, 201–209, 1949.

(10) C. T. Rajagopal and A. Venkataraman, 'The sine and cosine power series in Hindu mathematics', J.R.A.S.B. **15**, 1–13, 1949.

(11) C. T. Rajagopal and T. V. V. Aiyar, 'On the Hindu proof of Gregory's series', Scr. Math. **17**, 65–74, 1951.

(12) C. T. Rajagopal and T. V. V. Aiyar, 'A Hindu approximation to Pi', Scr. Math. **18**. 25–30, 1952.

(13) R. C. Gupta, 'Second order interpolation in Indian mathematics up to the fifteenth century', Ind. Jour. Hist. Sc., **4**, 86–98, 1969.

(14) R. C. Gupta, 'An Indian form of third order Taylor series approximation of the sine', Hist. Mathematica, **1**, 687–89, 1974.

(15) C. T. Rajagopal and M. S. Rangachari, 'On an untapped source of medieval Keralese mathematics', Arch. for Hist. of Ex. Sc. **18**, 89–101, 1978.

(16) B. B. Datta and A. N. Singh (Rev. by K. S. Shukla), 'Use of calculus in Hindu mathematics', Ind. Jour. Hist. Sc., **19**, 95–104, 1984.

(17) B. B. Datta and A. N. Singh (Rev. by K. S. Shukla), 'Use of series in India', Ind. Jour. Hist. Sc., **28**, 103–129, 1984.

(18) D. Pingree, 'Power series in medieval Indian trigonometry', in P. Gaeffke and D. A. Utz Eds., *Science and Technology in India*, Vol. II, University of Pennsylvania, Philadelphia 1985, pp. 25–30.

(19) C. T. Rajagopal and M. S. Rangachari, 'On medieval Kerala mathematics', Arch. for Hist. Sc., **35(2)**, 91–99, 1986.

(20) T. Hayashi, T. Kusuba and M. Yano, 'Indian values of π derived from Āryabhaṭa's value', Historia Scientiarum, **37**, 1–16, 1989.

(21) T. Hayashi, T. Kusuba and M. Yano, 'The correction of the Mādhava series for the circumference of a circle', Centauras, **33**, 149–174, 1990.

(22) Ranjan Roy, 'The discovery of the series formula for π by Leibniz', Gregory and Nīlakaṇṭha, Math. Mag. **63**, 291–306, 1990.

(23) D. Gold and D. Pingree, 'A hitherto unknown Sanskrit work concerning Mādhava's derivation of the power series for the sine and the cosine', Hist. Scientiarum, **42**, 49–65, 1991.

(24) K. Balagangadharan, 'Mathematical analysis in medieval Kerala', in *Science in the West and India*: *Some Historical Aspects*, Ed. by B. V. Subbrayappa and N. Mukunda, Himalaya Publishing House, New Delhi 1995, pp. 290–312.

(25) V. J. Katz, 'Ideas of calculus in Islam and India', Math. Mag. **68**, 163–174, 1995.

(26) T. Hayashi, 'Calculations of the surface of a sphere in India', Science and Engineering Review of Dohshisha University, **37.4**, 194–238, 1997.

(27) C. K. Raju, 'Computers, mathematics education, and the alternative epistemology of the calculus in the Yuktibhāṣā', Phil. East and West, **51**, 325–362, 2001.

(28) D. F. Almeida, J. K. John and A. Zadorozhnyy, 'Keralese mathematics: Its possible transmission to Europe and the consequential educational implications', J. Nat. Geo. **20**, 77–104, 2001.

(29) J. Bronkhorst, 'Pāṇini and Euclid', Jour. Ind. Phil. **29**, 43–80, 2001.

(30) D. Bressoud, 'Was calculus invented in India?', College Math. J. **33**, 2–13, 2002.

(31) Kim Plofker, 'Relations between approximations to the sine in Kerala mathematics', in G G Emch *et al.*, Eds. *Contibutions to the History of Indian mathematics*, Hindustan Book Agency, New Delhi 2005, pp. 135–152.

(32) P. P. Divakaran, 'The first textbook of calculus: Yuktibhāṣā', Jour. Ind. Phil. **35**, 417–443, 2007.

(33) D. Mumford, 'Why I am a Platonist', EMS Newsletter, Issue 70, Dec. 2008, pp. 27–30.

(34) D. F. Almeida and G. G. Joseph, 'Kerala mathematics and its possible transmission to Europe', in P. Ernest, B. Greer and B. Sriraman, *Critical issues in Mathematics Education*, University of Montana, 2009, pp. 171–188.

(35) D. Mumford, 'Review of "Mathematics in India" by Kim Plofker', Notices of the AMS, **57**, 385–390, 2010.

Some Overviews of History of Calculus

(1) C. B. Boyer, *The History of the Calculus and its Conceptual Development*, Dover, New York 1949.

(2) M. E. Baron, *The Origins of the Infinitesimal Calculus*, Pergamon, Oxford 1969.

(3) C. H. Edwards, *The Historical Development of the Calculus*, Springer, New York 1979.

(4) V. J. Katz, *A History of Mathematics: An Introduction*, 2nd Ed., Addison Wesley, Reading Mass. 1998.

K Ramasubramanian
Cell for Indian Science and Technology in Sanskrit
Department of Humanities and Social Sciences
IIT Bombay, Mumbai 400076
E-mail: kramas@iitb.ac.in and

M. D. Srinivas
Centre for Policy Studies
6, Baliah Avenue, Luz, Mylapore, Chennai 600 004
E-mail: mdsrinivas50@gmail.com

Notes on *Yuktibhāṣā*: Recursive Methods in Indian Mathematics

P. P. Divakaran

1. Overview

1.1. *Yuktibhāṣā*

It has taken a long time for historians of mathematics to move from curiosity to uncertain admiration to well-informed scholarly recognition of the brilliance of the mathematicians/astronomers who lived and worked in Kerala (on the southwest coast of India) from the last quarter of the 14th century CE until the end of the 16th. As the undertaking of bringing to critical attention the totality of their work gathers pace, one thing has already become clear: it can no longer be doubted that the high point of their mathematical achievements is the invention of calculus and its systematic and sophisticated development for application to trigonometric functions. Three personalities have emerged as key figures in this story. The first of course is the still shadowy Mādhavan (Mādhavan Emprāntiri, or Mādhava in Sanskritised form; throughout this article, I employ the Malayalam way of writing names from Kerala, with a terminal n), the founder of the school, who is credited by his followers with having created much of their strikingly original mathematics and whose creative genius pervades everything that they subsequently did. Then we have the polymath Nīlakaṇṭhan (Nīlakaṇṭha Somayāji), the pivotal link between Mādhavan and the later generations and the author of a large body of surviving work, notably *Tantrasaṃgraha* (*TS* from now on). And the third is Jyeṣṭhadevan who wrote what can accurately be called the first textbook of calculus, *Yuktibhāṣā* (*YB* from now on). *YB* is a comprehensive account, in its last two chapters, of the fundamental principles of integral and differential calculus (in that natural order in the Kerala approach to calculus), as well as the relationship between them, and their use in the study of trigonometric functions.

Until recently, *TS* has served as the *ur*text for students of the final flaring of the long tradition of Indian mathematics and astronomy that the Kerala school represents. It is a compendium of the new mathematics of Mādhavan and of its use in astronomy, in a compact Sanskrit verse format, and it is precisely dated (1500 CE)

at the midpoint of the relevant period. Within a relatively short period of about half a century after its writing, a detailed commentary appeared in two separate works by Śaṅkaran (Śaṅkara Vāriyar), *Yuktidīpikā* and *Laghuvivṛti*, providing details of several of the constructional and computational rules of *TS*.[1] *YB*, though it advertises itself in the opening sentence as following *TS*, is quite a different type of work. Rather than being a commentary (*vyākhyā* or *bhāṣya*), it is a stand-alone exposition of the material covered in *TS*, segregating (unlike *TS* itsef) the mathematics in a first part and its use in the construction of an accurate planetary model in the second. The mathematical Part I is all that will concern us here and what the abbreviation *YB* will refer to. It is distinctive in other ways too: at a time when all scientific texts were composed in Sanskrit, often in terse verse, it was written in Malayalam prose which is anything but terse, it is self-contained, and it provides reasoned justifications (*yukti*, which I will use interchangeably with 'proof' from now on)[2] for all results cited. In other words, *YB* gives us a portrait of the mathematical culture of its time as few other Indian texts of any period do. And since it also cites all earlier developments relevant to its concerns, it is an ideal vantage point from which to survey the entire landscape of Indian mathematical thought of which the work of the Kerala school forms the final summit. An English translation of *YB* by the late K. V. Sarma has recently been published [4], enabling scholars without Malayalam to finally get to know it almost at first hand[3]. The present article has taken advantage of its publication by assuming that the reader will have it to hand; I shall cite it frequently as [Sarma], along with the relevant section numbers, even though the translations of passages that often accompany the citations are my own.

---

[1] All three texts have been critically edited and published in the original Sanskrit in one volume by K. V. Sarma (with an introduction in English) [1]. For detailed accounts in English of some of the questions that *TS* is concerned with, see individual articles in [2].

[2] This is not to be taken to mean that the two terms refer to identical processes, either logically or, given the very different approaches in the Indian and European traditions to the enterprise of acquiring and validating knowledge, epistemologically. It will not be misleading to think of *yukti* as proof by other logical means than the now-universally practised sequence of steps starting with axioms. This point has been the subject of much impassioned discussion recently. The reader will find a detached evaluation of the two standpoints in the writings of Roddam Narasimha, especially [3].

[3] For those who read Malayalam, there is the admirable and invaluable edition (of Part I) of Rama Varma Tampuran and Akhilesvara Ayyar [5] (cited often in the rest of this article as [TA]) with diagrams, annotations and explanations of the more obscure parts of the text, now 60 years old and out of print; every subsequent piece of writing on *YB*, without exception, owes a debt of gratitude to this meticulous work. In English, Sarasvati Amma's book [6] has an excellent account of the details of the calculus work of the Kerala school (among many other topics), largely faithful. Sarma has chosen to title his translation as *Gaṇita-Yukti-Bhāṣā*, a choice whose motivation is something of a puzzle – almost all contemporary references to the work and all known copies, whether on palm leaf or, relatively recently, on paper, use the simple title *Yuktibhāṣā*. This can be doubly confusing since there is a Sanskrit work under the name *Gaṇitayuktibhāṣā* which, according to Sarma himself (in the introduction to his own translation of *YB*), is a second rate rendering of the Malayalam work by an incompetent.

Indian mathematical reasoning is as striking in its avoidance of certain techniques, for instance *reductio ad absurdum* methods, as for its partiality to certain others. This article is about what we can learn from *YB* and its recapitulations of earlier material about one such widely favoured general procedure, that of recursion. Three broadly distinct manifestations of the general idea of recursion (for the sense(s) in which I use this term, see below, section **1.7**)[4] find significant use in *YB*: recursive description, recursive computation and construction and recursive proof. The term recursive description does duty also for what in modern logic will be called recursive definition – as is well known, Indian thought, and certainly Indian mathematics, did not much care for *ab initio* definitions. Moreover, it is not an infrequent experience while going through *YB* to come across recursive characterisations of mathematical objects, the coefficients in a power series for example, arising in the course of a computation. This, and the fact that Indian proofs are largely constructive, make it not always easy to disentangle description, computation/construction and proof in a given piece of writing.

Recursive ideas run through Indian mathematics from before the time of Āryabhaṭa, notably in the perfecting of the decimal place-value notation for the natural numbers. But it is in the intricate reasoning that goes into the development of calculus that they finally find their most sophisticated expression, their natural home as it were. This has a reasonable explanation. Recursive processes are, generally speaking, nonterminating as is the infinitude of numbers; indeed, in the context of Sanskrit linguistics, it was recognised very early, presumably by Pāṇini (5th C. BCE) himself and certainly by his commentator Patañjali (2nd-1st C. BCE), that a science of language based on a finite set of recursive rules applied to a finite set of linguistic units has the capacity to describe its structure in all its unbounded richness. Patañjali's graphic way of illustrating this principle is worth quoting (see Staal [7], p. 40):

> Now if grammatical expressions are taught, must this be done by the recitation of each particular word for the understanding of grammatical expressions . . . ? No, says the author, this recitation of each particular word is not a means for the understanding of grammatical expressions. For we have a tradition which describes how Bṛhaspati addressed Indra during a thousand divine years going over the grammatical expressions by enumerating each particular word, and still did not attain the end. With Bṛhaspati as the professor, Indra as the student and a thousand divine years as the period of study, the end could not be attained . . . The recitation of each particular word, therefore, is not a means for the understanding of grammatical expressions. How, then, must grammatical expressions be understood? Some work containing general and particular rules must be composed.

In the realm of mathematics, the discipline of calculus cannot, by the nature of its very content, but deal with the infinite and its opposite, the infinitely small.

[4]Numbers in boldface refer to the sections and subsections of the present article.

It is no surprise then to find that *YB* accommodates and utilises infinitely iterated recursions in more than one role. But, as will become evident by the end of this article, not all of the recursive methods it calls upon are in response to the demands of the infinitesimal geometry which is the particular avatar that calculus takes in *YB*, nor do they all lead to infinite processes. What they do have in common is the constant presence of a certain spirit, that of the decimal place-value number system. It is for this reason that the next section **2** is devoted to a retracing of this frequently traversed path, ahead of the technically more demanding mathematical aspects in the subsequent sections. In the present section **1**, which is an overviw of the entire material, I have tried to avoid overly technical matters to the extent possible so as to give a reader who is not inclined to delve into mathematical details a general idea of the mode of thinking that animated Indian mathematics.

1.2. Gestures and Sounds, Ritual and Language

Going back as far as records and reliable tradition allow, three easily identified landmarks in the increasingly productive uses of recursive methods in Indian thought can be noted. The first began, probably before the earliest vedic works were composed, with the rules that govern all vedic rituals and effectively culminated in Pāṇini's systematisation of the grammar and syntax of the Sanskrit of his time.[5] An extensive ritual such as a *yāga* performed over several days is built up from its component elements in several stages of organisation and sequencing, the whole process controlled by strict rules which specify how subunits of the ritual, at various levels of organisation, may be combined, repeated, substituted, inserted one into another, etc., i.e., how they may be operated upon. That the number of repetitions of an operation is also specified we may take to be a performative limitation, not one of principle. Very similar rule-based structures are imposed also on the articulation of sounds (*mantra*) that accompany the gestures and actions of the ritual. It would seem then that a ritual is fundamentally a 'syntactically' defined object, perhaps serving certain functional ends, its 'semantic' signifiance being secondary; in other words, we are talking about "rules (largely) without meaning".[6] One may even argue, more generally, that divorcing sense from symbol bestows freedom; that attaching meaning to each component part of a recursively constructed object or process, as opposed to an overall purpose, might well hinder the realisation, through a process of abstraction, of the full power and flexibility afforded by rules without meaning. Nothing illustrates the power of rule-based abstraction more dramatically than modern mathematcs.

---

[5]It will be evident to the informed reader that my understanding of Indian rule-based thinking and its impact on mathematics has been strongly influenced by the work of Frits Staal.

[6]A faithful and comprehensive record of an actual performance of a complex ritual is available in Staal's magnificent work *Agni* [8].

As for Pāṇini, the quotation above from Patañjali says it all, or virtually all, about the indispensability of having a finite set of rules which may be applied over and over again without bound as the only means of imposing structure and regularity on a natural language. It needs to be said that Pāṇini's principles cover all aspects of Sanskrit as a language-in-use, their power and depth already evident at the elementary level of the small collection of *sandhi* rules which describe/prescribe the way words (primarily their sounds; meaning is hardly relevant) are combined and in the process transformed. It is possible, though not of great significance for us, that these and other rules were a conscious adaptation, from the ritual syntax that apparently pays little heed to meaning, to the setting of a natural language which cannot possibly dispense with meaning at a fairly low level of organisation, that of a sensible word. The other difference is that, as Patañjali's fable highlights, a language does not have any finiteness limitation.

1.3. One Thing after Another . . .

The second landmark is without doubt the adoption of the place-value system with 10 as base for the naming, writing and manipulation of integral numbers. It is fairly certain that this entailed a long drawn out evolution, beginning with the first mentions of number names in the earliest texts[7] and arriving at a mature, economical and consistent model for numbers by the time of Brahmagupta (7th C. CE). Along the way, all the features that mark what the Arab mathematicians called the Hindu numbers became incorporated: the role of zero both as a numeral, the one preceding 1, and as the cardinality of the empty set as in the place-value notation; the systematic ordering and naming of numbers going up to values beyond any practical utility, making the idea of the infinite almost palpable as one might say; fractions and the rules for operating with them; negative numbers and their arithmetic properties, and so on. It is this finished product that was transmitted to Europe from the Arab world, beginning in the 12th-13th century, by Leonardo of Pisa (Fibonacci) and possibly others. The fascination that the 'discovery of the zero' has held for European historians has perhaps tended to turn our attention away from the role of these other elements in making the decimal system the universally adopted basis for human numeracy. Things are beginning to change now. (For the early history of the zero, see Staal [10] and for the impact that a proper understanding and use of negatives had on Indian arithmetic and geometry, see Mumford [11], both in this volume).

---

[7]A succinct historical account will be found in the book of Datta and Singh [9] (Vol. I). Citing instances from the *Ṛgveda* in the well known *Vedic Index* of Macdonell and Keith, it says on page 9: "We can definitely say that from the very earliest known times, ten has formed the basis of numeration in India".

It is often assumed that the idea of a place-value number system came to India from the Babylonian sexagesimal numbers. The reasons for the belief are, firstly, that the sexagesimal system had attained a degree of completion by about 1800 BCE, some centuries before the first vedic mentions of numbers in the *Ṛgveda*. Then, the decimal notation itself is foreshadowed in the Babylonian use of a single character for 10 in the writing of sexagesimal numbers. Finally there is the prevalence in Indian astronomy of successive divisions by 60 in measuring time as well as the circumference of the circle. There may well be an element of validity in each of these reasons but, given our poor knowledge of the relevant history, none can be taken to be a decisive argument. For instance, the astronomical use of 60 can only be dated back to Āryabhaṭa (born 476 CE) who, on the time scale of interest here, is a late figure. It is perhaps not an accident that the advent of 60 in the study of the (celestial) sphere and the measurement of time coincides with the sudden blooming of astronomy associated with him, plausibly through contact with the Alexandria of Ptolemy, and had no influence on a pre-existing decimal system – nothing non-astronomical has ever been counted in units of 60 in India. The prudent conclusion has to be that we know too little about the historical factors bearing on the issue to take a position, keeping in mind the risk of mistaking causality (the principle that says that the past cannot be influenced by, but *may* influence, the present) for causation. I shall return to the question of the uncertain history later (section **2.4**).

Whatever its antecedents, the decimal system is the first purely mathematical expression of the power of recursive construction in the Indian context. Its significance for all mathematics, indeed for virtually all human activity, and for all time to come, can only be described as seminal. Even though it took some four centuries after its introduction by Fibonacci for it to definitively displace the laborious methods of doing arithmetic practised by European savants, the speed of its subsequent spread was rapid. By the end of the 18th century, Laplace was writing enthusiastically about it, as in the oft-quoted tribute: "The ingenious method of expressing every possible number using a set of ten symbols (each symbol having a place value and an absolute value) $\cdots$ seems so simple nowadays that its significance and profound importance are no longer appreciated" and, further, "The idea of expressing all quantities by nine figures whereby is imparted to them both a position and an absolute value is so simple that this very simplicity is the reason for our not being sufficiently aware how much admiration it deserves".[8] An even more significant tribute to its power, not so much in the eloquence of his words as in the remarkable use he put it to, came

---

[8]Post-revolutionary France officially adopted (in 1799) decimally ordered (metric) units for length and mass mainly as a result of the advocacy of Laplace (along with Lagrange). The sexagesimal division of time was obviously too well entrenched to be meddled with. The opposite happened in India much earlier: the adoption of the new sexagesimal time divisions in the 5th century CE barely touched the decimally organised counting numbers dating back to the *Ṛgveda* or the possibly equally ancient but less systematic multiples and subdivisions of lengths and weights.

a century earlier from Newton. Understanding that use, in calculations with power series, needs the context to be prepared and I will come to that in a while.

What the decimal system first does in practice is to take a notion of a (countable) quantity, which humans apparently have an innate faculty for when the numbers are small but becomes uncertain and hard to manage as they increase, and give it the same capacity for precision no matter how large the numbers are. That this problem was a preoccupation is clear from a number of early vedic passages. Some of them refer to situations in which numbers are matched or paired off without their magnitudes being specified, implying an understanding of when two numbers are equal even if each is not known 'absolutely', which is a plausible starting point in trying to quantify a discrete set. Then we have the verse from the *Ṛgveda* quoted by Staal in [12] and first highlighted by Renou, about the fire god Agni being the only being capable of apprehending precisely, we are not told how, large numbers. Over a period of time, but metaphorically at one stroke, the new way of counting gave people (and gods, not counting, naturally, Agni), the means to deal with arbitrarily large numbers with the same facility with which they dealt with small numbers. (That decimal counting did not completely supplant comparison counting at one stroke may be the message of the well-known passage from the *Yajurveda* in which a list of powers of 10 goes hand in hand with a pairing off of cows and sacrificial bricks; for the full passage, see for instance [13], p.131). One might say, from a modern viewpoint, that it replaces an abstract recursive *definition* – no one does any arithmetic starting from the Dedekind-Peano axioms – by a recursive *constructive procedure* of limitless power. Not the least significant aspect of this taming of numbers is that it goes hand in hand with their naming, by resorting to a relatively small collection of simple proper names for the atomic numbers[9] 1 to 9 as well as for 10 and its powers, supplemented, ideally, by two clearly differentiated rules for combining these names, one for addition to and the other for multiplication of powers of 10. And to name is to bring into being, to make exist. The wonder of this achievement is still alive in the pages of *YB*: in the very first section, after quoting a verse from Bhaskara II on the names of powers of 10 upto *parārddham* (10^{17}), there occurs the statement, "If we endow numbers with multiplication and positional variation (*sthānabhedam*), there is no end to the names of numbers; hence we cannot know [all] the numbers themselves and their order" [Sarma 1.4][10]. The rest of the opening chapter is then devoted to a concise but complete account of the basic arithmetical operations on decimally written numbers.

---

[9]More generally, in a place-value system with base b, I use this term for the numbers 0 to $b-1$, mindful of the association of the term 'digit' with the base 10.

[10]My translation; Sarma's version deviates apparently slightly but, for our purpose, crucially. Throughout this article, all unattributed translations are mine. The Sanskrit-Malayalam words and phrases cited are spelt as they are in *YB* and in Malayalam generally.

1.4. The Infinite and the Infinitely Small; Calculus

Georges Ifrah's book [14] has a table giving the names of powers of 10 which goes up to 10^{421}. While it is difficult to be extremely precise about the antiquity of the very many lists of names of powers of 10 made by Hindus, Buddhists and Jainas alike (see for instance [9] or Hayashi's book on the Bakhshali manuscript [15]), all of them probably antedate the 3rd century CE or so[11] by which time sectarian Buddhist texts were being written in prolific quantity, many containing cosmogonic speculations or stories about the omniscience of the Buddha involving large, enormously large, numbers. (If anything, the Jainas were even more number-happy). For a while, it would appear that such semi-quantitative ideas of the very large coexisted with qualitative notions of the infinite as expressed in the oft-quoted upaniṣadic passage concerning the concept of *pūrṇam* (full or complete) as the quality of being unchanged when something is added or taken away, or in the early Buddhist canonical references to the Buddha's agnosticism about the spatial and temporal boundedness or otherwise of the universe. The number lists are just a witness to the dawning realisation that there was now a bridge to infinity, constructed span by decimal span; a bridge without end, but one that would carry them as far as they wished to go.

By the time of Bhaskara II (12th century CE), Indian mathematicians, who were also astronomers most of them, no longer felt the need to glory in their mastery of astronomical numbers. The list of numbers in his *Līlāvati* stops at *parārdha* (which is now 10^{17}) and for *YB*, 10^{17} is a perfectly acceptable point at which to halt in the interminable journey and do the rest of it in the imagination: *YB* leaves no doubt that it understands the resort to a fixed large number, as a substitute for an infinitely large one, to be a matter of practical convenience and ease of explanation[12]

At the other extreme, the infinitely small caused no difficulty at all, once the approach to infinity itself was brought under control: as a number was taken to larger

[11]Some of the more modest lists are in Hindu texts that are very much earlier – the *Yajurveda* has a list [9] going up to *parārdha* which, if the lists name all the consecutive powers of 10, stood at that time for 10^{13}. There are in fact many lists of varying antiquity and provenance [14, 15] and they do not all agree on all the names.

[12]The first 'main theorem' of the calculus part of *YB* is the infinite series for π. The derivation of the series proceeds, as is well known now, by dividing half the side of the circumscribing square of the unit circle (= $\tan \pi/4$) into n segments and letting n tend to infinity (see further down in this subsection). In the account of this procedure, there occurs a passage which says in part, "The smaller the segment, the more accurate the result. Therefore imagine that each side is cut into atoms and then carry out the summation. For this, if the division is by *parārddham*,. . . ., also divide by *parārddham*". It then goes on to say that to get the exact value of the integral (*saṃkalitam*), the division has to be endlessly fine. For a translation of the full passage, see the section entitled "Integrals of Powers" of my article [16] or section 6.4.1 of [Sarma].

and larger values, its reciprocal came closer and closer to zero. In a general sense, division (at least by a non-zero number) and the concomitant fractional numbers seem to have posed no conceptual problems in India from an early time, just as subtraction and negative numbers did not. Even in *YB*, the operation of division is dismissed in a very brief paragraph after an explanation in detail of several different ways of carrying out multiplication. Similarly, the chapter on fractions is only concerned with the mundane computational aspects of extending the basic operations on integers to fractions – mostly about finding a common denominator. When these operations are subsequently applied to geometric quantities, we see the same pragmatism in play. A segment of arbitrarily small size of a straight line or an arc was arrived at simply by cutting it into a number n of equal pieces and then letting n increase without bound. We might say that the only limit that explicitly occurs is the limit $n \to \infty$, never the Newtonian limit $x \to 0$; no philosophical energy needed to be spent on defining or describing *ab initio* the notion of an infinitesimal. The contrast with the European story of calculus, where 'local' questions such as the problem of tangents (finding a sequence of chords to a curve that approached the tangent at a point) and of extrema (determining the points at which a function or its curve attained its local maxima and minima) were from the beginning in focus, is clear enough to need no further comment.

The derivation of the series for π is a good illustration of this approach.[13] The problem is that of finding the length of (or rectifying) an arc of the circle (of unit radius) equal to one-eighth of its circumference, i.e., subtending an angle $\theta = \pi/4$ at the centre, and it is in its solution that the numerical series for π and its generalisation, the arctangent series, (described later in section **3**) first saw the light of day. Schematically, the method consists in cutting up the unit tangent to the circle at any point ($\tan \pi/4 = 1$) into n equal parts, approximating the arc-segement corresponding to each tangent-segment by its chord (more accurately, by half the chord of twice the arc, a little twist that may justly be considered the founding step of Āryabhaṭa's trigonometry), adding them up and finally letting n grow large. Right at the start of the description, there is the remark, "The greater the number [i.e., n], the more accurate the circumference". (*YB* treats the case of the general arctangent series later by means of an elementary property of similar triangles; here I do not distinguish between the general and the particular cases). So it is clear that the procedure was effectively understood to be what we will consider today a specially simple instance of the use of the fundamental theorem of calculus: to determine a function explicitly, first differentiate it and then integrate the derivative. But the procedure did not, as it did not later for Gregory and Leibniz,

[13]Most of the rest of this section is, unavoidably, about the actual mathematical content of *YB*. Though I have kept equations at bay, it will be useful to have in hand an account of the mathematics as given for instance in the supplement of explanatory notes [17] by Ramasubramanian, Srinivas and Sriram (abbreviated from now on to [RSS]) in [Sarma]; better still, come back here after reading the rest of the present article.

give a closed expression for the circumference. A variety of recursive tools are now put to work. First, each term in the sum over the segments, before the large n limit is taken, is itself turned into an infinite series by means of an elementary recursive identity, a result which continues to hold in the limit. (The recursive identity will get its fair share of attention in the following sections). Though there is nothing said about the possibility of the series not converging at this point (it is dealt with subsequently), the student/reader is left in no doubt that the sum of the whole series is the exact length of the arc which one may then approximate by truncation at any point. In this step, the debt that the free use of the (actually) infinite series owes to the (potentially) infinite series of powers of 10 which are the decimal representations of integers appears direct and undeniable. Over a century after the writing of *YB*, the story repeated itself in Europe in a different guise when Nicolaus Mercator and Isaac Newton first wrote down the logarithmic series and the binomial series. Newton felt the need to provide a rationale for his formal manipulation of infinite series from "the doctrine recently established [!] for decimal numbers" in a revealing passage from *De Methodis Serierum et Fluxionum* written in 1670-1671 (translated by Whiteside [18], vol. III, p.32) and it merits being quoted *in extenso*:

> Since the operations of computing in numbers and with variables are closely similar – indeed there appears to be no difference between them except in the characters by which quantities are denoted, definitely in the one case, indefinitely so in the latter – I am amazed that it has occurred to no one (if you except N. Mercator with his quadrature of the hyperboia) to fit the doctrine recently established for decimal numbers in similar fashion to variables, especially since the way is then open to more striking consequences. For since this doctrine in species has the same relationship to Algebra that the doctrine in decimal numbers has to Arithmetic, its operations of Addition, Subtraction, Multiplication, Division and Root-extraction may easily be learnt from the latter's provided the readers to be skilled in each, both Arithmetic and Algebra, and appreciate the correspondence between decimal numbers and algebraic terms continued to infinity; namely that to each single place in a decimal sequence decreasing continually to the right there corresponds a unique term in a variable array ordered according to the sequence of the dimensions [meaning the powers[14] of 10, 'places'] of numbers or denominators [presumably, in fractions in decimal notation; Simon Stevin had already decimalised fractions] continued in uniform progression to infinity (as you will see done in the sequel).[15]

[14] I am grateful to David Mumford who reminds me that this usage of the term 'dimension' derives from Greek geometry in which squares and cubes of lines referred to 2- and 3-dimensional spaces – Newton just extended the terminology to numbers and to higher powers.

[15] For the structural reasons that make this more than just an analogy, see my article [16]. At the time of its writing I was ignorant that Newton had actually been inspired by the decimal system in his use of power series just as was, I had suggested, Mādhavan.

For Newton this was no mere contingent justification. Six years later, in the earlier of the two letters to Oldenburg for forwarding to Leibniz, the *Epistola Prior* ([19] Vol.II, p.32]), he says:

> Fractions are reduced to infinite series by division; and radical quantities by extraction of the roots, by carrying out these operations in the symbols just as they are commonly carried out in decimal numbers.

What Newton had in mind clearly is that a decimally represented integer can be thought of as the value of a polynomial when the 'variable' is fixed at 10: if c_i is the entry in the ith place of a number n, then n is the value of $\sum_i c_i x^i$ at $x = 10$ – one has of course to keep in mind that the entries in a decimal number are restricted to the digits 0 to 9 and so the correspondence requires the carry over rules to be imposed on them when two numbers are added or multiplied. Further down (section **5.1**), we shall see how just this correspondence is utilised by *YB* to model the algebra of polynomials in one unknown (an arbitrary positive odd integer) on decimal arithmetic.

1.5. A Recursive Proof

The use of the recursive identity mentioned above is not an essential part of calculus *per se*. It is an auxiliary or enabling tool in much the same way that binomial series were for Newton. (Special cases of the infinite series generated by the identity in fact coincide with special cases of Newton's binomial series, one specific instance being the present case of the arctangent series). Its use leads to a representation of each term in the series as a finite sum which in turn will become an integral, the integral of a positive even power, in the limit of large n. These integrals are still to be evaluated and the next key step forward, that of a recursive proof, makes its first appearance here.

In a daring step announced with little fanfare,[16] *YB* recognises that the way to evaluate the even integrals is i) to consider the integrals of all nonnegative powers, both even and odd, and ii) to relate the value of the integral of a given (kth) power to that of the $(k-1)$th power, i.e., an even (odd) term is obtained from the preceding odd (even) term. This is an astonishingly original thing to have done. Nothing that went before in Indian mathematics prepares us for this step, which should not be surprising since what the step encodes is the principle of integrating by parts or, equivalently, the rule for differentiating products. It

---

[16] *YB* has a very matter-of-fact and pedantic style; none of the light-hearted asides favoured by other writers, such as addressing questions and explanations to the slow-witted or the adept – Bhāskara II poses one problem to a " proud expert in algebra" (*bīje paṭutābhimānaḥ*) – can be found in its pages.

is also a conscious application of the principle of induction, in itself a (if not the most) sophisticated manifestation of the power of recursive methods. What makes it work is that the recursive step is generic: the procedure that generates the kth term from the $(k-1)$th term does not depend on k. The first step, that of considering *all* k, *no matter how large*, is thus forced by the desire for an exact answer and the resulting recursive nature of the proof. Or, perhaps, it is the other way: the impulse is the recursive mindset and it is that which makes an exact answer attainable.

All of this is covered in chapter 6 of *YB*, which then concludes by treating a practical problem raised by (convergent) infinite recursion: how can one extract useful numbers from a procedure that is endless? If the successive terms in an infinite series diminish in value rapidly, one can, optimistically, be satisfied with truncating the series after a finite (small) number of terms, the number depending on the rate of convergence and on the accuracy demanded. The arctangent series converges notoriously slowly. *YB* has some quite remarkable ways of accelerating the convergence by a process of reordering the terms, achieved by the use of a method of recursive correction. It also poses the question of how one may estimate the remainder left out by a truncation at the ith term, i.e., the difference between the exact value (the sum of all the infinite number of terms) and the sum up to i terms. The impressive way this is done, no longer to our surprise by now, is again recursive in content but with an interesting modification. A first guess is made about its value, which is then fed into the equations satisfied by the remainder to generate a 'truncation error' whose value (*sthaulyam*, grossness) is used as a guide in fine-tuning the initial guess rather than as the input in the next iterative step. The reason is that this would-be recursion within a recursion becomes impossibly complicated to implement after the first stage. The step that is iterated is the one of going back to the first guess and changing it judiciously. Already in two or three such iterations very precise values for π are produced.

1.6. Solving a Differential Equation

The bulk of chapter 7 of *YB* is devoted to a problem that goes back to Āryabhaṭa, that of expressing the sine of an arbitray angle in terms of that angle or, in geometric language, given an arc, to find an expression for half the chord of twice the arc. In a famous (semantically nonsensical) verse (*Gītikāpāda* 10 of *Āryabhaṭīya*, abbreviated to *AB*), Āryabhaṭa sets out the results of his computation in the form of a table of sines for 24 equally separated angles in the first quadrant, i.e., in 24 steps of $\pi/48$. He himself in an opaquely worded stanza (*Gaṇitapāda* 12 of *AB*) and his immediate successors in more transparent words, *Sūryasiddhānta* for example, have indicated the procedure used to make the table. With help from later works, in particular *YB*, the principle behind the construction can be described qualitatively as follows.

When the angle is close to zero, the value of its sine is close to the value of the angle if angles are measured, as is commonly done today, in natural units, namely radians. Equivalently, both the angle and the sine can be expressed in sexagesimal minutes as Āryabhaṭa did. (This was a nontrivial matter; Ptolemy's astronomical calculations were made exceedingly involved by his partial use of degrees and minutes). To the accuracy that Āryabhaṭa worked to, which was a minute of arc, $\pi/48$ was small enough for this to be true as could be verified by easy computation, for example starting with $\pi/6$ and halving it repeatedly three times. Thus the first two entries in the sine table were filled. The next step required a notion of sine differences, namely the difference between the sine of a given angle in the table and its predecessor (this was what Āryabhaṭa actually tabulated in his verse), and then, of their second differences, namely the difference between two successive such sine differences. Elementary circle-geometry then leads to an equation for the second difference. The recipé outlined in *Sūryasiddhānta* is most simply interpreted as arising from an approximate recursive solution of this difference equation, terminated after the first iteration, which is accurate to the required degree (or, one should rather say, to the minute), an interpretation well supported by the relevant portion of *YB* (section 7.4.2 of [Sarma]).[17] I am conscious of the inadequacy of this summary of one of the cleverest parts of early Indian mathematics but, once again, invite the reader to go to the technically complete account in section **4.2**.

The journey from Āryabhaṭa's small angles and sine differences to the infinitely small angles and sine *differentials* of *YB* took close to a thousand years and also took some false turns.[18] But the recognition that the equality of an angle and its sine

[17]A clear and early (1795) exposition in English of this material is that of the Scottish geometer and historian of geometry, John Playfair [20] (He of the parallel axiom but also, more pertinently for us, one who closely followed the work of H. T. Colebrooke on Indian science. Charles Whish, the first to write in English on some of the Kerala texts including *YB*, was also strongly influenced by Colebrooke). Only in one respect is Playfair slightly off the mark: he conjectures that "the brahmins" had in hand the geometrical equivalent (Ptolemy's theorem etc.) of the addition formulae for the sine and the cosine. That is unlikely; what they had was what they needed, the formulae for small increments (see section **4.2**). I am grateful to Roddam Narasimha for the Playfair reference and for making me appreciate its historical importance in the reconstruction of Āryabhaṭa's sine table.

[18]*Gaṇitapāda* 11 of *AB* is a strong indication that Āryabhaṭa himself understood perfectly well that his procedure was an approximate one and that it could be made more accurate if desired by a finer division of the circumference. But, until we get to Kerala, his followers seem to have led themselves astray. In his *bhāṣya* of *AB*, Bhāskara I asks the intriguing question: does a circle have an arc equal to its own chord? and answers: yes, otherwise an iron ball resting on the floor will not be stable; such an arc will be (1/100)th (a fraction smaller than (1/96)th!) part of a circle (see [21] for instance). More interestingly, Bhāskara II arrived at the correct formulae for the surface area and volume of a sphere by dividing a great circle into 96 equal parts and effectively equating $2\pi/96 = \pi/48$ to its sine (see Sarasvati Amma's book [6] for the details). We have to conclude, on available evidence, that little conceptual progress was made in the time intervening between Āryabhaṭa's cryptically expressed realisation of the need for indefinitely continued subdivision (*Gaṇitapāda* 11) and Mādhavan's full-fledged calculus of the circle. I hope to return to the question of the prehistory of Indian calculus elsewhere.

is only a limiting property of angles tending to zero, presumably by Mādhavan, and its being put to use to transform Āryabhaṭa's approximate table into Mādhavan's exact series seem to have been almost simultaneous. There are two parts to this transformation. First, Āryabhaṭa's equation for the second differences was shown to become exact as the step size was taken towards zero becoming, automatically as it were, the second order differential equation satisfied by the sine (and the cosine) function – this is calculus pure and simple, exactly as we might proceed today. But the recursive solution of the equation now has to be carried through all the way, through an infinite number of iterations, to get an *exact* expression for the sine; that is precisely what *YB* does, resulting in the famous series in powers of the angle.

There are several extremely ingenious steps involved in this progress from Āryabhaṭa (finite differences and approximate numerical solutions) to Mādhavan (differentials and exact analytic solutions), some of them wonderfully illustrative of the flexibility of recursive techniques – most remarkable is an instance of a doubly inductive proof which will get its deservedly full treatment later in section **4.4**. But none of that should blind us to the revolutionary new way of doing mathematics that made it possible in the first place, the way of calculus that Mādhavan pioneered.

1.7. Recursion and Rules of Composition

The following sections aim to provide the mathematical details of some of the more striking uses to which recursive methods are put in *YB* (as well as some historical background where relevant and available), after first taking a closer look at that model recursive object, the natural numbers as expressed in the place-value notation. The approach will be almost entirely descriptive. There will be nothing here relating to the elaborate theoretical framework(s) constructed around the general idea of recursion in linguistics, logic, foundations of mathematics, computational theory and so on, the main reason for the reticence being lack of knowledge and competence. But it is not an unreasonable stance to adopt when dealing with Indian mathematical thought which famously (or notoriously, depending on one's epistemological leanings) declined to take categorical positions in regard to issues of definitions, axioms, rules of deduction, etc., etc. It is not that Indian men of learning were indifferent to the reliability or otherwise of their ways of establishing what is true – we have only to read Nīlakaṇṭhan to know that methods of acquiring and validating knowledge, both empirical and theoretical, were a constant concern among, at least, astronomers and mathematicians. Nothing was self-evident to these inheritors of a tradition, going back two millennia or longer to the time of the later *upaniṣad*s and the *śrāmaṇa* movement, of questioning and sceptical enquiry. Their position appears to have been that an excessive preoccupation with first principles, if there be such, was a distraction from the more important business of understanding

the world as it is.[19] They might have been intrigued but not surprised if they could have known that present-day mainstream scientific practice, after having explored the axiomatic methodology to its limits, has gradually been turning towards a more easy and pragmatic view of such foundational questions; that these are issues the practitioners leave to others to worry about or, at worst, think about on Friday afternoons.

Having said this, this very spirit of realism should persuade us to try and evaluate the place of recursive methods, in the very broad sense in which we take them, as a productive means of creating interesting new mathematical knowledge. In his work on the structure of rituals (and, inevitably, its relation to the structure of language), Frits Staal ([7], p.42) characterises a recursive rule as one which "can be applied to its own output" and so "can produce infinitely many forms". Intuitively, this is an appealing definition, general enough to embrace the instances that we are going to be concerned with. But when dealing with a developed mathematical science from our current perspective, it will need to be both sharpened and broadened; the emphasis on infinity, for instance, will have to be relaxed. It can well (and does) happen that, for any chosen initial input, the output after a certain finite number of steps coincides with the initial input or the input/output of a preceding step. An exact recursive solution of the exact difference equation for the sine will be a case in point – if the incremental angle is $2\pi/N$ for example as it was for Āryabhaṭa, the output of the Nth step will be the same as the initial input because of the periodicity of the sine function. More particularly, such cyclic recursions will occur whenever the mathematical structure we are computing in is defined on a finite set (but finiteness is not a necessary condition). It would be pointless to go through the same cycle over and over again, learning nothing new, but it is not a serious matter in itself (except, perhaps, for a badly programmed computer which has not been told when to stop); the interest that a recursively generated result may hold is not necessarily in proportion to the number of steps needed to get to that result.

A more pertinent, broader, issue is that the manner of our response to modes of thinking from the past is necessarily subservient to a certain tyranny of progress. No historian can fully unlearn what he has learned of subsequent events, however conscious he may be of the imperative of placing himself in the epoch. In the case of recursive mathematical techniques, this psychological barrier is a particularly difficult one to penetrate. The reason is simply that an essential part of the progress that tyrannises has consisted in reorienting mathematics towards an abstract structural mode of thinking even when the particular problem at hand may arise in a concrete context. Such structures are defined completely by rules governing the operations

[19] In some of his more philosophical writings, Nīlakaṇṭhan gives us a clear idea of the epistemological moorings of his work and, presumably, that of others of the Kerala school; especially revealing are his remarks on the futility of seeking logical first causes. These writings have recently been critically examined by Roddam Narasimha in a series of articles which can be retraced from [3].

that can be performed on the elements of a given (structured) set, in particular rules of composition associating to *every* pair of elements precisely one element of the same set. Since the rule of composition is defined on every pair of the set, it follows that it can be applied to its own outputs over and over again. More generally, a problem of interest may involve a collection of such structurally defined sets with rules governing 'maps' from one set of the collection into another (including composition within one such set), which can themselves be successively applied subject to various conditions of compatibility. In short, recursive operations are embedded in the very genes of modern mathematics, their deployment routine and automatic; progress has turned the creative into the commonplace. But they do still catch the eye when a description or a proof depends critically and explicitly on the recursive properties of, say, composition rules. Instances are easy to come by. In the area of recursive proofs in particular, the very general technique of proof going under the name of mathematical induction is without doubt the prime example. Here let us only recall that inductive proofs proceed by establishing a correspondence between an infinite number of propositions and the recursively defined natural numbers.

The abstract structural vision that animates much of modern mathematics was of course not even a gleam in the eye of Mādhavan and the school he founded. Nevertheless, in assessing the special role that recursive reasoning played in their work, we have little choice but to use the same criterion that we would in a contemporary situation: how good is the mathematics they produced and could they have reached their mathematical destination by a path that was not the one they so consciously built, one step after another? Were there other more banal, 'natural', means available to them which would have done the job? The quality of the mathematics is by now self-evident. As for the method, I hope a reader who persists to the end will be convinced that the techniques they perfected and used so imaginatively and to such good effect were a natural continuation of the entire tradition of Indian mathematics and that there was no other way conceivable by them that would have taken them to their goal. It is in this perspective, as an account of the culmination of a long evolution, that this essay should be viewed.

2. The Grammar of Numbers

2.1. Seeing and Counting, Naming and Knowing

Measurement and quantification are an essential part of the means by which rational humans have sought comprehension of the world they live in. The ability to quantify is a direct prior factor in our understanding of 'nature', the world in all its aspects; the interrogations "how many?" and "how much?' precede "how?" and "why?". And counting is the first, the most primitive, measurement there is. All possible answers to "how many?" are in principle definite and absolute. They do not need to be compared with externally prescribed standards for them to make sense in our

own minds or to be communicated to others; no balance, meter stick or clock is required in the process of counting.

But this is a hard-won autonomy it would seem. By all accounts, early humans had no problem with counting small numbers – they 'took it in at a glance'. Staal [12] quotes Louis Renou to the effect that in the *Ṛgveda* the verbal root *khyā*, from which comes *saṃkhyā* meaning 'number', signifies 'to look' and that it is only in later vedic times that it acquired a numerical connotation. When it came to larger numbers, the early vedic people apparently (going by the cows-for-bricks passage) had to resort to comparison and correspondence, not with a fixed standard number, but with numbers chosen by the situation, a sort of 'weighing' of pairs of numbers in an abstract number-balance as it were.[20] That this was so is entirely in the nature of things; comparison is still the only way we have of quantifying magnitudes such as length or time which do not come in multiples of an irreducible, universal unit. If the Planck length had a human scale, the physical world would be discrete in the large and all measurements would consist of counting; vedic geometry certainly, and perhaps even Greek geometry, would not have had to be invented.

How large had a number to be for people to fail to 'see' it? From the historical evidence we have, we might guess, not beyond 10 – even the Babylonian sexagesimal system had a single separate symbol for 10.

We can take the passage from the *Ṛgveda* about Agni, " He has seen or counted (people) like one observes groups of cattle with an owner of cattle" (Staal's translation [12] from Renou's French) in two senses: that Agni could see 'at a glance' or count the number of people as he could the number of cattle, or that he could count while others merely saw. In either sense, Agni could count, presumably beyond ten. One must assume that, for those who could not, the capacity to match went with the capacity to distinguish between 'more' and 'less' – to draw a meaningful conclusion from observing which pan of the number-balance went down – and by how much, as long as the difference was, again, small. That would introduce a semiquantitative (because of the limitation to small numbers) notion of order among numbers but would still not bring a mortal the divine gift that Agni possessed. The way forward from here is the obvious recursive one of extending the ordering by repeated comparisons. It is difficult at this point not to jump ahead and note that it is this imagined primitive idea of order, limited in the beginning to nearby numbers and subsequently extended by recursion, that is captured by Peano in full quantitaive generality in the succession axiom. But although the Peano axioms give a logically complete characterisation of natural numbers, they still do not help directly either

[20]It is of course acknowledged that drawing conclusions about the evolution of ideas and practices from early vedic sources is a hazardous business simply because of the strong possibility of a long time-gap between their actual composition and their consolidation in the *saṃhita*. Along with examples of primitive, comparison-based, enumeration the *Ṛgveda*, for instance, also has a rich repertoire of decimally designated number names.

to take in large numbers 'intuitively' or, in practical terms, to do arithmetic; to implement this last step we are once again obliged to go back to the primitive idea of succession and build everything up from the operation of adding 1 at a time (see section **2.3** below).

"There is no end to the names of numbers; hence we cannot know the numbers themselves and their order". To know numbers and to use them, we have first to give them names, an identity that will assign each of them its exact place in the order of all numbers and that will let us know instantly what that exact place is. For it to serve its purpose the nomenclature has to be rule-based – in order to let mere mortals see where a number denoted by a name stands in the order – and economical – so as not to tax their finite powers of memory. We might well paraphrase Patañjali and say, "How, then, must the meanings of numbers be understood? Some work containing general and particular rules must be composed". We might go further and formally prescribe the "general" rules as the rules designating a number N by the unique ordered sequence of atomic numbers (including 0) $n_0, n_1, \ldots$ which constitute its place-value representation, namely the coefficients of the powers of the chosen base b, not necessarily 10, in the sum

$$N = n_0 \times (b^0 = 1) + n_1 \times (b^1 = b) + n_2 \times (b^2 = b \times b) + \ldots,$$

expressed orally or in writing or both. In case the expression is verbal they need to include also the rules by which the number is given a verbal identity, which is accomplished by a set of names for atomic numerals juxtaposed in the correct order by use of joining or compounding rules applied at both the semantic (*samāsa*) and the syllabic (*sandhi*) levels. (If the representation is written down in symbols as in Mesopotamia, names are strictly speaking not necessary, but those who wrote and read them probably talked to one another). Thus made tangible, numbers are the first, the most primitive, concept in Indian, perhaps all, mathematics.[21] "That which is the particular study of numbers relating to enumerables (*saṃkhyeyam*) is mathematics" (*YB*, introducing the place-value system and the rules of arithmetic). As for "particular" rules, the natural candidate is the collection of the laws of arithmetic worked out within the place-value representation, including the carry over or overflow rules. The particuar rules cannot be given in a concise or even practical way without the general rules having been established. How will one teach

---

[21] There is good evidence that numbers have priority over geometry in the vedas not only conceptually but also chronologically. The *Śulbasūtra* manuals containing the earliest accounts of geometry in India cannot be dated earlier than about 700 BCE while number names with a distinct decimal structure occur freely already in the *Ṛgveda*. For all matters concerning the chronology of vedic tmes, I have relied on the synoptic account of the current state of knowledge given in [13]. As for the earlier Harappan culture, we can say little definite about their numbers (see footnote 23 and section **2.4** below), but a fair picture of its geometric sophistication is now beginning to emerge [22]. It is, however, not ruled out that there may be a degree of continuity between Harappan geometry and that described in the *Śulba* texts.

or learn, for instance, the multiplication table in Roman numerals except, as with Bṛhaspati and Indra, by endless enumeration?

2.2. The Power of 10

One begins, then, by giving names to the atomic numbers 1 to 9 (in the decimal base case), *eka, dvi, ⋯, nava* and the base itself, 10, *daśa*. Ideally, 11, 12, . . . , can now be designated by an application of the joining rules: one-plus-ten (*ekādaśa*), two-plus-ten (*dvādaśa*) and so on till we get to 20 at which point, a different joining rule should come into play to create names for two-times-ten and similarly for 30 and so on. Combining these two joining rules, we have a system that, in principle, names all numbers up to 99 and it requires exactly ten distinct names and the convention of the two joining rules to be remembered. When it comes to 100, we have options. The first is to express it as the sum or product (or a combination of the two) of smaller numbers in very many different ways using the supposedly already acquired rules of arithmetic, an option that is available for numbers less than 100 as well but generally not used. Apart from the lack of uniqueness and the consequent absence of a repeatable structure in the naming rules, such an *ad hoc* system is bound to reach its limit sooner or later and new number names and more *ad hoc* rules will have to be introduced. The preferred Indian way was to give a new name to every power of 10 so that, with minimal demand on resources, the naming method that works up to 99 can be indefinitely extended. It is also a fully recursive approach to naming and, so, to counting – one starts with a finite ordered set of names for the lowest nine of the ordered set of all numbers and two linguistic rules of composition (the joining rules); the repeated application of those rules together with progressively less frequent new names for the powers of 10 generates the decimal nomenclature of every number.

That is the ideal. In practice the ideal was never fully realised. It is not in the nature of natural languages to be completely logical and Sanskrit, vedic or later, is no exception. In the event, there have always been deviations from the ideal in the way numbers were designated, most markedly in the proliferation of synonyms for the same number, often arrived at by decomposing it as sums and products in different ways.[22] At the same time, astronomers and mathematicians from Āryabhaṭa onwards were constantly on the lookout for a rule-based, symbolic but verbalisable representation that would not lead to very long names for big numbers and so not inconvenience the versifiers unduly. The goal was achieved

[22] The book of Datta and Singh [9] has an extensive discussion of number names and number symbols and of possible reasons for some of these deviations. They suggest in particular that the use of different numerical expressions for the same number as well as that of *bhūtasaṃkhyā*, which is an association of numbers with certain conventional multiplets (3 worlds, 8 directions, 9 planets (*graha*) and so on) has its origin in the exigencies of Sanskrit versification.

finally with the invention and widespread use of the ingenious *kaṭapayādi* notation, especially popular in Kerala. Its basis is a systematic one-to-many map of the atomic numerals into the consonant syllables that takes advantage of the organisation of most of the Sanskrit (and Malayalam) consonants in groups of five, the latter then being used in a place-value format, turning numbers into mnemonic 'words' which could generally be contrived so as to give them a sensible meaning. Thus a given number has many verbal representations, giving the system a welcome flexibility. What matters more is that the inverse map from a 'word' to a number is well defined and easy to figure out; in particular, the syllabic length is exactly the number of 'places' (examples: *mā-tu-laḥ* = 365, *dhī-ja-ga-nnū-pu-ram* = 210389, both from *YB*). It is also properly recursive and comes closest to the abstract ideal, namely a purely symbolic representation, without need or room in it for names for any number other than the atomic ones (including for powers of the base), even for oral communication – one might say that with the *kaṭapayādi*, Indian mathematics finally settled its dues to the Sanskrit syllabary. It is then something of a surprise to find it used so very sparingly in *YB*, written at a time and place where it was all the rage. Perhaps this was because *YB* was free from the metrical discipline of verse. What it prefers to use instead are the common Malayalam (and Tamil) number names, more systematic than Sanskrit-based ones and still in currency.[23]

It is true that many of the points made above are not directly relevant to a strictly symbolic written representation such as the Babylonian sexagesimal system. In order to read their numerical tablets, we do not need to know what names the Babylonians gave to their atomic numerals 1 to 59, leave alone what joining rules they had. We do not even need to know whether they had names for their numbers at all. It is nevertheless an amusing exercise to compare bases: what relative advantage or drawback can be discerned in the choice of a particular base, 10 or 60? Whether written or spoken, there is no escaping the requirement that each atomic numeral has to have an individual name or symbol which will have to be memorised. The Babylonian strategy of employing a base-within-a-base to simplify writing – clusters of the unit symbol up to 9 and then a new symbol for 10, similarly clustered for multiples of 10 – we may take as having been devised mainly to overcome the barrier of 10. (Did the names they must have had use a similar strategy?) Did the Indian idea of using 10 not merely as a notational abbreviation, but as the base in the place-value representation, emerge out of this and out of the realisation that it is wasteful to have two different fixed numbers playing two distinct roles where one would do? In our current state of historical knowledge, the question has no

[23] There is one singular feature of these names worth mentioning: the number 9 is designated as '1 less than 10', a usage which spills into the names of 90 and 900 (though not, as in Sanskrit-based languages, into 19, 29, etc.). Taken together with the identity of the word for counting and the name of the number 8 in Dravidian languages, it has given rise to speculation that the old Dravidian civilisation counted in base 8; see Fairservis [23], quoting Kamil Zvelebil, who make the case that this may actually be a legacy from the Indus valley language.

satisfactory answer. It is also probably moot; the very emphasis on naming points to a more likely source of inspiration, the proximate one of the phonetic organisation in the vedic period of the Sanskrit syllabary. It may in fact very well be that both the number names and the syllabary have a parallel early vedic origin and evolution.

In the Indian oral system, one had to have names also for the powers of 10; how many were needed depended on how high one wanted to count. It is completely elementary to see that, to count up to *parārddham* (10^{17}), one needs 27 (10 for the atomic numbers and 17 for the powers of 10) distinct decimal names, to be combined by the joining rules, and 69 ($60 + 9$, $60^9 < 10^{17} < 60^{10}$) sexagesimal names; to count up to 10^7, the corresponding numbers are 17 and 64 respectively. The decimal system wins simply because of its far smaller number of atomic numbers. This may appear a frivolous reason to choose 10 over 60, but in a dominantly oral culture, it is not to be dismissed as irrelevant. And when it comes to doing arithmetic, the advantages are more concrete. For instance, once the place-value multiplication procedure is put in place, a major achievement in itself, one has only to learn 'by heart' the multiplication table up to 9×9, something we all did in primary school, and then just go by the rules. The basic sexagesimal 59×59 multiplication table cannot ordinarily be memorised by mortals. As is well known, the Babylonians overcame this obstacle by using formulae like

$$m \times n = \frac{(m+n)^2 - (m-n)^2}{4}$$

and preparing written master-tables of squares.

2.3. Arithmetic

The fundamental arithmetical operation in *YB* is the addition of 1 together with its inverse, the subtraction of 1. All the eight operations it explicitly describes, general addition and subtraction, multiplication and division, taking squares and cubes, and taking square roots and cube roots, are built out of these two basic ones by, it is needless to say, their repeated application: "Addition is of use in multiplication, multiplication in squaring, squaring in cubing. Similarly, subtraction is of use in division, division in [taking] square roots, square roots in [taking] cube roots. Thus, the preceding [operations] will be of use in the succeeding" ([Sarma] 1.3). (The eight operations are first classified as those which increase and those which diminish). Addition of two numbers m and n is, as expected, explained as the n-fold repeated addition of 1 to m (in more modern words, repeated applications of the successor function) and similarly for subtraction. There is no explicit reference to negative numbers in this (first) chapter; what to do with $n - m$ for $n < m$ does not get a mention.

Multiplication of m and n is then defined, also as expected, as the addition of m to itself n times. ("Then multiplication – it is just addition if we reflect [on it]"). The only slightly surprising point arises in relating division to subtraction, accomplished by falling back on geometry in the time-honoured way. Imagine the dividend n to be the area of a rectangle one of whose sides is the divisor m. The quotient is then the maximum number of rows or columns, i.e., $m \times 1$ rectangles, that can be removed from the whole figure: $n/m := l$ such that $n - l \times (m \times 1) = 0$. Again, it is tacitly understood that n has to be a multiple of m; fractions are not mentioned here.

All of this is totally in conformity with the Indian approach to the logic of arithmetic from much earlier times. In particular, there is no sign of any sympathy for abstraction as might have been expected 400 years after Bhāskara wrote *Bījagaṇita*. Even by implication, there is no hint that subtraction of and division by arbitray positive integers will automatically force the extension of the natural numbers to include negative integers and fractions. Fractions in their own right make their entrance soon enough, in chapter 3, where, for example, division by n/m is carefully shown to be the same as multiplication by m/n. But we have to go deep into the calculus of the series for π to see the first algebraic statement about negatives, a citation of Brahmagupta's well known aphorism and its generalisation to indeterminate numbers ("quantities (*rāśi*) whose numerical values (*saṃkhya*) are unknown"). The terms in the π series are given by a recursive rule and they alternate in sign, so this belated recall of well-established principles and their easy adaptation to variables was inescapable. It is certain that there was no fear of negatives, there had not been any for a long time [11], but they were still, to facilitate computation, handled with care ("Here, if an indeterminate number (*rāśi*) is of negative nature, it must be distinguished by some mark"; the context is computation in indeterminates with the aid of tokens, [Sarma] 6.8, p. 75).

It is equally clear that there was, if not fear, a degree of wariness regarding irrational numbers, even after Mādhavan had broken through the rationality barrier with his π series, the significance of which was well understood by Nīlakaṇṭhan. The suspicion of irrationals also goes back a long way (see [16]). In *YB* itself, the procedure explicitly described for extracting square and cube roots of numbers presupposes that the answers will be whole numbers. Later, in the chapter on fractions, it is explained that $\sqrt{n/m}$ is $\sqrt{n}/\sqrt{m}$, but n and m are still squares. Throughout the history of Indian mathematics, there does not appear to have been any attempt made to understand the essential difference between fractions and numbers such $\sqrt{2}$ (or π for that matter) till we get to Nīlakaṇṭhan, even though during and after Āryabhaṭa's time the irrational π was often approximated by the irrational $\sqrt{10}$ rather than by more accurate and easier-to-handle rational approximations. Nor, apparently, was the possible irrationality $\sqrt{2}$ ever broached despite accurate algorithms for finding rational approximations to it being known from at least the time of the *Śulbasūtra*.

The fear of the irrational may provide an indirect explanation for another enduring mystery: the absence of a decimal representation for fractions in Indian mathematical work. It would have been a natural and undemanding thing to have done, as demonstrated convincingly by the Babylonians. Can it be that the avoidance of decimal fractions has something to do with the fact that even simple fractions can have decimal expressions that are actually, not just potentially as in the case of whole numbers, nonterminating?

There is an irony here. When it came to geometry, there was no way to keep away from square roots since one of the two pillars of Indian geometry, the theorem of the diagonal (Pythagoras' theorem; the properties of similar triangles form the other pillar) cannot possibly dispense with square roots. In the early part of chapter 6, *YB* describes the (infinitely iterative) procedure of getting at the circumference from the diameter by starting with a circumscribing square and repeatedly doubling the number of sides through circumscribing regular polygons. (It is a good guess that Āryabhaṭa found his approximation to π in this way). The procedure is of course riddled with square roots. It must have rankled. Before starting on the first steps that will finally lead to the π series, *YB* announces: "After this, [I] describe a way of producing the circumference for any desired diameter without having to take square roots". With such undramatic words were his pupils initiated into the new-born discipline of calculus by Jyeṣṭhadevan.

2.4. Origins?

Let us conclude this bird's-eye view of the grammar of numbers with a brief return to the question of origins and possible influences. The natural, in fact the only, mathematical candidate that might have served as a model for the decimal place-value 'language' – it *was* a language, orally expressed, with its own grammatical rules – is the Babylonian sexagesimal system.[24] The mathematical apogee of Mesopotamia, roughly 2000-1600 BCE, was contemporaneous not with the vedic but with the earlier Indus valley or Harappan civilisation in its late phase. Some vestiges of west Asian civilisations have been found in Indus valley sites, but they are so sparse and inconclusive as to be useless for our purpose. The reverse flow is somewhat better authenticated, in the form of a number of seals, sealings, etc., unearthed in an

[24]The possibility of a Chinese influence is definitely discounted by the profusion of number names reflecting an evolved decimal structure in the *Ṛgveda*, including in its earliest *maṇḍala*s, a fact already alluded to several times. (A forthcoming article in collaboration with Bhagyashree Bavare will look at the evidence in detail). The earliest Indian references to China are related to commercial activities and do not predate the 3rd century BCE (in the *Arthaśāstra*). Long before that time, not only a perfected decimal counting system but also fairly advanced arithmetical procedures using it (the *Śulbasūtra*s of Baudhāyana and Āpastamba) were well in place. For a quick *tour d'horizon* of Sino-Indian intellectual engagement, which flourished most vigorously in the first millennium CE, see Amartya Sen's essay "China and India" in [24].

area spreading from the Arabian coast through the mouth of the Tigris-Euphrates into modern Syria. As for the Indus culture within its native domains, the ruins and the excavated artefacts are witness to a high degree of geometrical and metrical sophistication, extending from the construction of their cities and the varied symmetric patterns on a fairly extensve sample of pottery, seals etc. [22], to the very accurate regularity of the weights. The latter in particular persuade us to conclude that it was a predominantly binary culture (or one based on a low power of 2), the weights doubling in steps from 1 to 32 units, with weak indications of a decimal progression as they got heavier. Then there are the signs showing the supposed numerals 1 to 7 by vertical strokes, giving rise to the speculation already cited (see footnote 23) that counting was octal.[25] All in all, there is little that indicates a special role for the number 60 in the greater Harappan culture in its evolved form over its entire duration (of at least a millennium) and extent (from modern Baluchistan to the Indus river basin, to the river Yamuna to the east and across Saurashtra). Nor is there any such evidence for the vedic culture[26] that succeeded it in northwest India and then spread quickly into most of the Ganga river basin.

But it is fair to say that Harappan research is still in an early consolidation phase. Despite the vigour that animates the field, going beyond the material remains to a deeper insight into the culture has proved to be an elusive quest; we are nowhere near getting inside the minds of the Indus people and we will not get much closer as long as their writing remains unread. One can always hope that there will be a breakthrough on the decipherment front one of these days and that the writing will then turn out to contain information about the west Asian civilisations with which they had close commercial ties. While waiting for that to happen, what is one to do? The course that recommends itself is to concentrate on whatever we can learn from what we do have. There is no question that our understanding of the vedic intellectual legacy is much richer and deeper compared to the sketchy and uncertain picture we have, or are likely to have in the near future, of the Indus culture in its less material aspects. That rich vein is far from exhausted. From what we know about vedic Sanskrit and Pāṇini's grammar and from what we can learn about the Indian approach to numbers from even as late a text as *YB*, it does not seem to be a far-fetched idea to look for links between the two.[27] Indeed, already at the level of phonology, there are striking similarities. Sanskrit organises its syllables

[25] I thank Nisha Yadav for a valuable update on the current state of knowledge on the Indus civilisation. It should be added that the idea of an octal base has also had some doubters.

[26] There are sporadic mentions of the number 60, as there are of other multiples of 10, in the *Ṛgveda* but no indication that the powers of 60 held any special significance (work in collaboration with Bhagyashree Bavare) as would be expected if 60 was used as a base.

[27] The case for doing so goes back to the vedic studies of Louis Renou and more recently has been most persuasively made by Frits Staal over a period of time; for an account of Staal's ideas, see [12]. The whole of the present section 2 can be thought of as a limited excursion in that direction.

in a well-ordered sequence, the vowels first and then the pure consonants in five groups of five each (plus a few outliers which don't fit the pattern), the order of the groups depending on the part of the vocalising organs utilised in articulating them, progressively from the back of the palate to the lips, see for instance [25]. (There is also a well-defined order within each group). The genesis of this organisation goes back to vedic times, its final systematisation coinciding roughly with the era of Pāṇini [25]. It is easy to see in the pure syllables an analogue of the atomic numerals; there are rules for combining them to produce all legitimate sounds of the language which are in turn used to formulate rules to combine clusters of sounds or 'words'. In the context of phonology, the vowels and the simple consonants are 'as if' *prakṛti* (see footnote 28 below), all other sounds being construced out of them. (The ordering of the syllables and the combination rules are of importance in the *kaṭapayādi* mapping of numbers into verbal expressions; for instance, the vowel value of a syllable has no numerical significance, just like all but the last sound of a compound consonant).

Perhaps, rather than think of Sanskrit grammar as a possible imperfect model for the Indian number system, we should open our minds to the possibility that both are parallel manifestations of a particular intellectual mindset, specific to the time and the place. Perhaps there was a mathematical Pāṇini, perhaps he even wrote a book entitled *Daśādhyāyi*, alas lost forever.

3. Varieties of Recursion – the Arctangent Series

3.1. Generalities, the Notion of *Saṃskāram*

The "doctrine of decimal numbers" which gave confidence to Newton in the handling of infinite series in his early work in calculus had been a constant presence in Indian mathematics from at least the vedic era. When the Kerala mathematicians found their own route to calculus, two and a half centuries before Newton, they were the inheritors of a decimal culture that was seeded two and a half millennia earlier and had grown to healthy maturity in the following thousand years or so. Though it is not in the nature of Indian mathematical writing to spend much time on motivation and history, the influence of decimal numerical thinking is so evident in all of it that it is commonly characterised as fundamentally computational. So it is not only in the Kerala school's fearlessness of infinite series (and of the idea of infinity itself) but in the totality of their mathematics that we come upon the pervasive undercurrent of recursive construction that is at the core of the place-value representation of numbers.

Recursive construction generally figures in Kerala writing on calculus in the guise of a *saṃskāram* (refining; occasionally, other synonymous terms are also employed). The result of a process of *saṃskāram* is something that is *saṃskṛtam*, that which is or has been refined. This is also the name of the language in which

the writing itself was done, Sanskrit, as distinct from *prākṛtam* or Prakrit, that which is natural.[28] There are variations in the details of how mathematical refining is carried out in different situations, but the general idea in its simplest variant can be summarised as follows. In a given problem (evaluating a geometrically defined quantity, solving an algebraic or differential equation, etc.), i) use insight and ingenuity to make a first guess of the answer, generally incorrect; ii) use the formulation of the problem itself (i.e., in ignorance of the correct answer) to obtain an exact expression for the difference between the exact answer and the first guess that displays the (unknown) exact answer in a practically usable form; and iii) repeat the process by substituting for the exact answer from the expression obtained in step ii), which will result in an updated guess displaying, again, the unknown answer, and so on. Stated like this the procedure may appear both pointless and vague. Slightly less vaguely, suppose the first guess for the unknown answer x is x_0 and the difference $x - x_0$ is a known and easily computable function f of x. Then $x = x_0 + f(x) = x_0 + f(x_0 + f(x)) = x_0 + f(x_0 + f(x_0 + f(x))) = \cdots$ (in practice it may be necessary to approximate f in some clever way). The examples from *YB* that we are going to be dealing with will illustrate the variations that are possible within this general framework.

But, before that, we may already take note of some of the qualitative features that emerge from these examples and illustrate how they vary from instance to instance. First, the procedure may be difficult to carry through beyond one or two iterations and so may have to be terminated, resulting in an approximation, often characterised as gross (*sthūlam*) or fine (*sūkṣmam*), to the correct answer. In one case where this happens (in the problem of estimating the truncation error in the π series), *YB* goes back and plays around with the initial guess so as to get a better approximation. The second possibility is that the procedure leads to the exact answer either in a finite number of steps (finite recursion, the most sophisticated example being *kuṭṭākāram* or *kuṭṭaka*, 'pulveristaion')[29] or by leading back on itself (cyclic recursion, as in

---

[28] *YB* is of course an exception, for reasons unknown. But it is so clearly in the Sanskrit *śāstra* mainstream that the actual language of writing can sometimes seem almost an irrelevance. Early in the first chapter, there occurs the elliptic sentence, "The numbers one to ten [written as words in the text] are like (as if they were) *prakṛti* [that which occurs naturally, the original]", followed by an explanation of how at every stage of extension from what I have called atomic numbers to successively higher powers of 10, the earlier stage of construction becomes the *prakṛti* for the next one. Tampuran and Ayyar [5] provide an enlightening footnote on the general connotation of the term *prakṛti* in the *śāstra* literature and the reason why *YB* has to resort to the equivocation of "as if". Malayalam grammar uses the same term to denote an 'original' word from which derivatives are constructed for related linguistic units such as case endings for nouns. It is useful to remember that the two hundred years during which the Kerala school flourished was also the time that saw Malayalam, having absorbed the massive verbal and structural invasion of Sanskrit, beginning to settle down to its modern identity.

[29] The method of *kuṭṭaka* was devised in its most elementary form by Āryabhaṭa to solve equations of the form $by - ax = \pm c$ in positive integers for positive integer coefficients a, b, c. The problem arises in matching (for example) the solar and lunar periods in astronomical time-keeping and has a

the putative exact solution of Āryabhaṭa's difference equation for the sine). The third and final possibility is the most interesting: that the recursion does not end but has a sufficiently regular structure for it to be carried through *ad infinitum* (infinite recursion). The resulting series are the most celebrated fruits of the Kerala calculus, but the main emphasis in the present article will not be on the series *per se*, nor on the deeper methodological advance (the infinitesimal method) that led to them, but only on the recursive element.

To illustrate some of the issues specific to infinite recursion, it is enough to refer to the first two examples from *YB*, the π series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \cdots$$

and its generalisation to the arctangent series for an arbitrary angle θ lying between 0 and $\pi/4$,

$$\theta = \tan\theta - \frac{\tan^3\theta}{3} + \frac{\tan^5\theta}{5} - \cdots .$$

Infinite series such as these, arrived at by a formal recursive procedure, run the risk of not converging to any finite answer. Whether by insight or otherwise, the Kerala mathematicians invariably avoid this ever-present pitfall – it is trivially easy to produce recursively constructed series which do not converge, as the following subsection will illustrate. The question of convergence is a particularly sensitive one for the arctangent series since the variable $\tan\theta$ is not bounded. It is quite obvious from the paragraph from chapter 6 of *YB* explaining how to go from the π series to the arctangent series that its author was aware of this and knew what to do about it. The first sentence, "Here [I] explain how to find the arc [$R\theta$] corresponding to the *bhujājyā* [$R\sin\theta$] or *koṭijyā* [$R\cos\theta$], *whichever is smaller*" ([Sarma] 6.6), makes it clear that what is sought is not a formal expansion of θ in powers of $\tan\theta$ for the entire range of values of θ (it is enough to consider angles between 0 and $\pi/2$ which is what *YB* does) but rather a convergent expansion in the variable appropriate to a given subdomain of values of θ, i.e., that the $\tan\theta$ expansion is the correct convergent one for θ in the first octant, $\sin\theta < \cos\theta$. The last sentence of the paragraph confirms that when $\cos\theta < \sin\theta$, the appropriate variable is in fact $\cot\theta$. (The full paragraph is part of section 6.6, to be read with Fig. 6.6 of the explanatory notes, of [Sarma]). If this is not sufficiently convincing evidence of the concern that the series should make sense, we have the whole of the very elaborate

whole chapter (5) in *YB*. The solution is a recursive one, involving the use of the Euclidean algorithm in a numerical (integral) form. It is very well explained in several places, e.g., [9], the epilogue (B) of [Sarma] and most thoroughly, especially in its connection with continued fractions, in the appendix (in English) to [TA]. For this reason, and since it has no direct linkage to any other of the recursive tools employed by *YB*, I shall have nothing further to say on the subject here.

discussion at the end of chapter 6 (sections 6.8, 6.9 and 6.10 of [Sarma]) of methods of accelerating the rate of convergence and of estimating truncation errors which would make no sense unless one knew that the series *was* convergent.

There is also in *YB* an instance to illustrate the other side of this question as it were: Mādhavan's interpolation formulae for the sine and the cosine. They are arrived at by a novel variant of *saṃskāram* dealing with two quantities ($\sin(\theta + \delta) - \sin\theta$ and similarly for cos) simultaneously and so provide a very interesting example of recursive techniques. But, unlike in the case of the power series around 0 for the same functions, *YB* treats the application to interpolation briefly, not going beyond the second order correction and dismissing the subject with a terse remark on how to improve the accuracy of the approximation by two more orders – it does not refer to the possibility of repeating the process indefinitely. And that is a good thing because, if it did, the result would be an infinite series for $(\sin / \cos)(\theta + \delta)$ in powers of δ, a would-be Taylor series which is not the Taylor series because the coefficients are wrong.

The other issue, related to this but more philosophical than technical, is one which seems to have caused some confusion among scholars. It is this: assuming that infinite recursion gives a convergent series in a given problem, did the Kerala mathematicians themselves think of such results as a series of successive approximations only, or as results which are in principle exact when 'all' the terms are added ('in the limit' as we would say), but which for practical purposes will have to be terminated at some point? Several historians have subscribed to the first view in recent times, sometimes explicitly. Partly, the confusion is due to the excessive emphasis put on these series in the early writing in English (and, often, even today) on the Kerala work, to the detriment of the driving philosophy behind it, that of turning Āryabhaṭa's approximation methods into a technique for getting exact answers by going infinitesimal.[30] But a reading of *YB* leaves no room for doubt that this is a mistaken view. Chapters 6 and 7 are liberally strewn with remarks which either say specifically that the results expressed by the appropriate series are accurate or exact when the recursion is carried out fully, or which are meaningful only if the author knew that to be the case: a few instances will be found in sections 6.3.1, 6.3.2, 6.3.3 and 6.4.1 of [Sarma]. (Chapter 7 is less explicit, often harking back to chapter 6

[30]Given the incredulity with which Western mathematical-historical circles slowly became aware of the discovery, in a distant corner of a British colony, of the texts containing these series, this is to an extent understandable. But there was another factor contributing to this misreading, which is the unavailability until now of *YB* in a widely accessible language. The Sanskrit-reading historians of Indian mathematics have been obliged to fall back upon two texts which can fairly be called (in the best vedic tradition) *Yuktibhāṣāpratinidhi*, *YB*-substitutes, namely *Yuktidīpikā* and *Laghuvivṛti*, Śaṅkaran's *vyākhyās* of *TS*. Both are written in a rigid Sanskrit verse format that discourages any indulgence in revealing asides such as we meet in the more loosely written Malayalam prose of *YB*. Consequently, while they are invaluable source books on the intricate geometric constructions and computations that go into the work on the series, they open only a narrow window on the conceptual isssues that the development of calculus had to face.

where these issues have already been confronted and settled). The conclusion is reinforced by *YB*'s silence in an aparently similar context, the interpolation formula referred to above; it is as though the power series and the numerical series for π were recognised to be exact in the limit but the interpolation series would not be and was merely a means of approximation.[31] In fact, evidence for the awareness of the distinction between an exact state of affairs and approximations to it forced by the demands of practical utility is already present in the very beginning of *YB*, in connection with the infinitude of numbers and the place-value representation of any arbitrary (finite) number. The sentence I have quoted earlier on the impossibility of knowing and naming numbers is followed immediately by: "Therefore, for the purpose of making use of them (or working with them, the Sanskrit/Malayalam word used is *vyavahāram*; "them" means the infinitude of integers), think as follows. The first place is for the numbers from one to nine. Then the place of all these multiplied by ten [is] the second. Let that be to the left. The place of one, the place of ten, and so on, are their names. That is the nature of numbers" ([Sarma] 1.2). This is as good a characterisation of the distinction between the (recursively defined) axiomatic and the (recursively constructed) arithmetical view of natural numbers as any we can think of.

3.2. The π Series and the Arctangent Series as Rectification

There are now available in English at least two detailed and faithful annotations of the *yukti* as described in *YB* of these two series, Sarasvati Amma's book [6] which has been around for a while now and the very recent explanatory notes [RSS] appended to Sarma's translation. With the assurance that the reader can fall back on these references for the details not discussed here, we can focus on just the recursive elements in the contents of chapter 6 of *YB* in which these series are exhaustively treated. They are abundantly supplied with diagrams, generally based on the Tampuran-Ayyar critical edition, and that is also a help: on the occasions on which we need to look at the specifics of the geometry, I can simply refer the reader to the appropriate figures from [RSS] rather than reproduce them once again. Also, since the order in which *YB* obtains these results – first the numerical series for π which is then generalised by a use of the properties of similar triangles – has no special interest from the recursive viewpoint, it is unnecessary (and often inconvenient) to stick to that order.

These series are answers to what in Europe came to be called rectification problems: one is required to find the length of a curve given either by an equation in

[31] Because of the historical interest of the misunderstandings about this aspect, it will be worth returning to it after a fuller discussion of the interpolation formulae (section **4.3**). Śaṅkaran's commentaries on *TS* give much greater space to these formulae than does *YB* and that may have contributed to magnifying their impact.

the Cartesian way or as the trajectory of a point under a prescribed motion. (In the present case, the curve is the circle of unit radius). The general approach used in *YB* to get to the answer is a precurser of what came to be called Riemann integration: divide the relevant interval of the x-axis (the independent variable, in our case $\tan\theta$) into n equal[32] segments, estimate the length of the curve corresponding to one segment of x by linearising it locally (here, by approximating the corresponding segment of arc by half the chord of twice the arc, i.e., by replacing the angle δ_i subtended by the ith segment at the centre by $\sin\delta_i$), add up the n contributions and, finally, let n increase without bound ('in the limit $n \to \infty$'). The limit makes the linear approximation exact, $\sin\delta_i \to \delta_i$ as $\delta_i \to 0$ and, moreover, allows the systematic neglect of terms of order $1/n$ in the sum as n is taken to infinity. The linearisation is of course the heart of the matter, the critical conceptual step beyond Ārybhaṭa's finitistic method. There are only two ways in which the procedure differs from what we teach today in a course in calculus, one minor and the other not so. The minor point is that the modern positive x-axis is, in the Kerala (and Indian) convention, the direction east which, in a diagram on sand spread on the floor, points away (and in the very rare diagrams on palm leaf, up), and the first quadrant is the E-S quadrant – so reflecting a Kerala circle through the N-S axis produces our conventional graph of the equation of the circle $x^2 + y^2 = 1$ (see Figs. 6.3 and 6.5 of [RSS]). The more interesting difference is that the work is carried forward as far as it will go before the $n \to \infty$ limit is taken, resulting in the uncovering of some striking, essentially arithmetical/algebraic, identities. The simplest of them is a finite counterpart of the formula for integration by parts (a simple special case of the Abel resummation formula, see section **3.4** below), a key step in the inductive evaluation of the integrals of positive integral powers. I draw these obvious parallels in order to remove any residual misgivings about the calculus credentials of *YB* and to highlight the fact that the postponement of the limit as late as possible is the essential difference between the Indian and the European approaches to calculus.

Having divided the unit tangent into n equal segments by points A_i (so the length $A_{i-1}A_i = 1/n$, independent of i), the geometry introduces the diagonals (*kaṛṇṇam*) connecting the centre of the circle to each A_i, of length d_i (Fig. 6.5 of [RSS]). The main geometrical result is that the sum of all the half-chords, $S_n := \sum_i \sin\delta_i$, is given by (the half-chords are denoted by b_i in [RSS]):

$$S_n = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{d_{i-1}d_i}.$$

[32]The equality is not essential for the procedure to go through. It is just the easiest way to avoid the direct introduction of the infinitesimal limit in which each segment is taken individually to zero. With the division by n, the only limit that needs to be taken is $n \to \infty$.

This is an exact result valid for any n. In the first instance of the use of the $n \to \infty$ limit, *YB* replaces $1/d_{i-1}d_i$ by $(1/2)(1/d_{i-1}^2 + 1/d_i^2)$. This is one of the very few steps for which *YB* provides no justification, but it is easy to see that the two quantities are equal to the first order in $(d_{i-1} - d_i)/d_{i-1}$. Thus S_n is approximated, first by

$$S'_n = \frac{1}{2n} \sum_{i=0}^{n} \left(\frac{1}{d_{i-1}^2} + \frac{1}{d_i^2} \right),$$

with $d_0^2 = 1$ and $d_n^2 = 2$ (normalised radius), and then by

$$S''_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_i^2}.$$

S'_n differs from S''_n by

$$S''_n - S'_n = \frac{1}{2n} \left(\frac{1}{d_n^2} - \frac{1}{d_0^2} \right) = -\frac{1}{4n}$$

which is ignored as $n \to \infty$. This last step is given a justification in the words, "As the segment of the side [the tangent] becomes small[er], the one-fourth part $[-1/4n]$ can be discarded" ([Sarma] 6.3.2), leaving no room for doubt about *YB*'s appreciation of the power of the limiting operation: S''_n will tend to an eighth part of the circumference ($= \pi/4$) and its expression in terms of the diagonals will tend to the π series.

3.3. The Elementary Recursive Identity

It is in expanding each term in S''_n in an infinite series that the recursive identity is first met with. It begins with the trivial equation

$$\frac{1}{a} = \frac{1}{b} - \left(\frac{1}{b} - \frac{1}{a} \right)$$

where a and b are, in the context, geometrically defined positive real numbers, namely $a = d_i^2$ whose reciprocal is the 'unknown' quantity to be computed in terms of the 'known' quantity $b = (\text{radius})^2 (= 1)$; thus $a = b$ is the first guess. The exact correction is

$$\frac{1}{a} - \frac{1}{b} = -\frac{a-b}{b} \cdot \frac{1}{a},$$

displaying the unknown a in a form suitable for iteration. The first *saṃskāram*, obtained by substituting the first guess in the correction term, gives the approximate

value

$$\frac{1}{a} = \frac{1}{b} - \frac{a-b}{b} \cdot \frac{1}{b}.$$

By repeated substitutions, we get

$$\frac{1}{a} = \frac{1}{b} - \frac{a-b}{b}\left(\frac{1}{b} - \frac{a-b}{b} \cdot \frac{1}{a}\right)$$

and so on, or, generally,

$$\frac{1}{a} = \frac{1}{b} - \frac{a-b}{b}\left(\frac{1}{b} - \frac{a-b}{b}\left(\frac{1}{b} - \cdots\right)\right).$$

The recursive identity is thus

$$\frac{1}{a} = \frac{1}{b} - \frac{a-b}{b^2} + \frac{(a-b)^2}{b^3} - \cdots + (-1)^k \frac{(a-b)^k}{b^k} \cdot \frac{1}{a},$$

valid exactly for all $k = 1, 2, \cdots$. The kth *saṃskāram* will consist simply of putting $a = b$ in the last factor of the last term.

It is obvious that the utility of the procedure depends on what kind of a handle we can get on the quantity $a - b$. In the problem of the π series, $a = d_i^2$ is the square of the hypotenuse and $b = 1$ the square of a side of a right triangle. So $a - b$ is the square of the third side, $a - b = i^2/n^2$; the kth recursive identity for $1/d_i^2$ then takes the form

$$\frac{1}{d_i^2} = 1 - \frac{i^2}{n^2} + \frac{i^4}{n^4} - \cdots + (-1)^k \frac{i^{2k}}{n^{2k}} \cdot \frac{1}{d_i^2}.$$

The name given to the terms (all but the last, one imagines; the terminology in the text is not completely unambiguous) on the right of the identity is *śodhyaphalam*. The process of determining the correction terms is itself not called *saṃskāram* presumably because that word is reserved for an approximation; and there is no intention here to stop the iteration after a certain number of steps and call it a day.[33] *YB* has other ideas, as comes across emphatically in the closing lines of the

[33] The literal meaning of *śodhyaphalam*, 'the result of refining (or purification)', is not very different from *saṃskāram*. One of the minor vexations of reading *YB* is what one may call the curse of the synonym. The availability of several different phrases for the same or a similar mathematical process seems occasionally to encourage the use of different, generally but not always synonymous, expressions for the same process as well as the use of the same expression for closely related but not identical processes. This is not unknown in modern mathematical writing, but can be a source of confusion in a mathematical language that dispenses with symbolic notation. One instance concerns the words *saṃkalitam* and *yogam* (see [16] for a brief discussion of this and other examples). In most cases, as in

paragraph: "In this way, if [we] divide all [the terms] by the multiplier ((radius)^2 (= 1 in our normalisation); i.e., if $1/d_i^2$ on the right is systematically replaced by the whole of the right side), the sequence of corrections (*śodhyaparampara*) will not end until [we] divide by the divisor (i.e., d_i^2). If [we] do not divide by the divisor, the sequence of corrections will not end. [All we can do is to] discard it (the correction term) when it becomes very small" ([Sarma] 6.3.3).

Apart from a precise formulation of how an infinite series is recursively generated – it will serve as a template for more complicated series later – there is enough in this passage to clarify several issues. First is the virtue of displaying the unknown answer in a way enabling repeated substitutive recursion *ad infinitum*. The demands of practical utility (*vyavahāram*) are met by the suggestion to terminate the sequence when it is deemed sufficiently accurate. But what stands out most is the dramatically repeated message that there is no need to end the sequence, that the exact measure of the circumference, in other words the value of π, can only be given by an infinite series. The message is reinforced in the opening lines of the following paragraph (see section 6.3.4 of [Sarma]) describing the even and odd terms with their alternating signs whose sum, in the limit $n \to \infty$, will become the exact circumference.[34] And it is difficult to read this passage without calling to mind how Nīlakanṭhan interpreted Āryabhaṭa's description of his own fractional value for π as *āsanna* (approximate), that it really means that π is irrational.[35]

All this leads directly to the question of convergence, particularly easy to deal with for the arctangent series because the recursive identity, infinitely iterated, results in a geometric series:

$$\frac{1}{a} = \frac{1}{b} - \frac{a-b}{b^2} + \frac{(a-b)^2}{b^3} - \cdots$$

with $-(a-b)/b$ as the common ratio. So, while the identity is trivially true at every stage of iteration for any values of a and b, the infinite series converges if and only if $a - b < b$ (in the present instance $i < n$ which is guaranteed; what does not work in the second octant is the geometric construction, as made clear by *YB*). Let us recall that it is in dealing with similar situations requiring the integration of functions which are not powers that Newton took recourse to formal binomial expansions. The binomial series for the inverse first power, which too is a geometric series, coincides with the iterative series, but a binomial expansion for fractional or general negative integral exponents seems to have had no use for the Kerala mathematicians. But

the present one, the context helps to remove any ambiguity. The Malayalam-to-English glossary at the end of [TA] translates *śodhyaphalam* as "correction to be applied to a result".

[34]It is this kind of insight that we are deprived of in the Sanskrit verse texts, see footnote 30.

[35]I have suggested elsewhere [16] that the irrationality of π may have acted as an immediate motivator for the development of the π series.

if the need had arisen, they probably had other effective techniques to hand, for instance a form of iterative approximation that appears to have been used already in the *Śulbasūtras* and subsequently in the Bakhshali manuscript to extract square roots. There are interesting recursive issues relating to these formulae which I take up briefly in section **4.6** where we will see that the *Śulbasūtra* formula may well be the first recorded instance of the use, in an arithmetical computation, of the general method of *saṃskāram*.

3.4. Recursion as Proof, Induction

Having shown that $S_n'' - S_n \to 0$ as $n \to \infty$, *YB* proceeds to demonstrate how $S = S'' := \lim_{n\to\infty} S_n''$ can be evaluated. This is done by taking the limit of each term in the sum

$$S_n'' = I_{n,0} + I_{n,2} + I_{n,4} + \cdots$$

with

$$I_{n,k} := \frac{1}{n^{k+1}} \sum_{i=1}^{n} i^k$$

for $k = 0, 2, \cdots$. The computation of $I_k := \lim_{n\to\infty} I_{n,k}$ in terms of which the circumference is finally expressed:

$$\frac{\pi}{4} = I_0 + I_2 + I_4 + \cdots$$

is explained in elaborate detail (sections 6.1 - 6.4 of [Sarma]) in what is probably one of the best organised pieces of exposition in the book. The contents are worthy of the care taken. Firstly, it encompasses the first results that can be considered 'pure calculus', answers to questions which had never been posed before, namely the values of the definite integrals

$$I_k = \int_0^1 x^k \, dx,$$

for all (not just the even) nonnegative integral k, formulating along the way the principle of integration by parts. (Their generalisation to an arbitrary upper limit is addressed later by means of a simple geometrical argument). More interestingly from the perspective of this article, the integrals are evaluated by a method which, except for the language, is indistinguishable from a modern inductive proof: the value of I_k is expressed in terms of I_{k-1} and so on till $I_0 = 1$ is reached. After

a prefatory sentence: "Even though it is not useful here, [I am] describing also integrals[36] of equals multiplied three, five, etc. times as they occur in the midst of those which are useful" ([Sarma] 6.4), *YB* first evaluates I_1 from $I_0(= 1)$ and I_2 from I_1 in detail, then I_3 from I_2 and I_4 from I_3 progressively more briefly, "by virtue of the reasoning given earlier". A final passage gives the general rule: "To make integrals of higher and higher powers, multiply the particular integral by the radius and remove from it itself divided by the number which is one greater (for unit radius, this amounts to $I_{k+1} = I_k - I_k/(k+2)$)" ([Sarma] 6.4.4). However, in order to highlight the recursive content, I compress all of this[37] into a description of the general case of evaluating I_{k+1} from I_k, following faithfully *YB*'s own steps in going from I_0 to I_1 to I_2.

The prior announcement about the need to consider odd values of k, "though not useful here", already offers a hint that it was an essential step in the proof and that the proof will turn out to be inductive. The details bear this out. True to its invariable practice, *YB* carries out as much of the procedure as is practicable for finite n, before taking the limit. In doing this it is a simplification to drop the powers of n from the definition of the integrals (*saṃkalitam*) and to focus on the sums of powers:

$$A_{n,k} := \sum_{i=1}^{n} i^k.$$

The first step in evaluating $A_{n,k}$ is none other than the first stage in a general process of *saṃskāram* as I have described it in section **3.1**. One makes an insightful first guess $B_{n,k}$ by replacing one factor of i in each term by n:

$$B_{n,k} := n \sum_{i=1}^{n} i^{k-1} = nA_{n,k-1}.$$

The 'error' is

$$B_{n,k} - A_{n,k} = \sum_{i=1}^{n-1} (n-i) i^{k-1}.$$

[36] The word employed is *saṃkalitam*, which is used in this part of the book for both $I_{n,k}$ and the limiting quantities I_k; see footnote 33.

[37] The thought naturally arises that Jyeṣṭhadevan was aware that this was an unprecedented sort of *yukti* and so was presenting his reasoning with the care and attention to detail that it called for. On the value placed by the Kerala masters on consensus as a means of validating a *yukti*, on convincing their peers and disciples of its correctness, see [3].

A simple reordering of the sum on the right side now leads to the identity

$$\sum_{i=1}^{n-1}(n-i)i^{k-1} = \sum_{i=1}^{n-1}\sum_{j=1}^{i} j^{k-1} = \sum_{i=1}^{n-1} A_{i,k-1},$$

so that the difference between the exact value and the first guess can be expressed as

$$B_{n,k} - A_{n,k} = \sum_{i=1}^{n-1} A_{i,k-1}.$$

So

$$A_{n,k} = nA_{n,k-1} - \sum_{i=1}^{n-1} A_{i,k-1}.$$

This is the key relation. In contrast to the application of the elementary recursive identity, however, $B_{n,k} = nA_{n,k-1}$ becomes a 'known quantity' only after the sum of the $(k-1)$th powers is computed. Moreover, the difference does not display $A_{n,k}$ itself in a nice form amenable to a simple iteration but only as a linear combination of sums of the $(k-1)$th powers of numbers upto $n-1$. A simple iterative computation of $A_{n,k}$ and hence of $I_{n,k}$ and I_k is therefore out of the question. But it opens the door for an inductive evaluation since it allows I_k to be related eventually to I_{k-1}. A second problem is that combinatorial complications make it difficult to handle $A_{n,k}$ themselves for any k but the lowest few.[38] So at this point, *YB* takes n to be asymptotically large before taking the inductive step, even for $k = 1, 2, 3$. Accordingly, it begins with $A_{n,0} = n \sim n$ ($\sim$ denotes the asymptotically dominant term) to show that $A_{n,1} \sim n^2/2$. We can save some writing, without in any way detracting from *YB*'s basic method, by short-circuiting its careful progress through $k = 2$ and $k = 3$ and dealing with the case of general k as a modern inductive proof would. Assume therefore that $A_{n,l} \sim n^{l+1}/(l+1)$ for all $l < k+1$. Then in the key relation for $A_{n,k+1}$, we have $B_{n,k+1} \sim n^{k+2}/(k+1)$ and, for the error term,

$$\sum_{i=1}^{n-1} A_{i,k} \sim \sum_{i=1}^{n-1} \frac{i^{k+1}}{k+1} = \frac{1}{k+1} A_{n-1,k+1}.$$

---

[38] $A_{n,0} = n$. The values of $A_{n,1}$, $A_{n,2}$ and $A_{n,3}$ were known since a long time. For general k, $A_{n,k}$ are expressible in terms of Bernoulli numbers, but there is no evidence that the Kerala school had exact expressions for them.

For $n \to \infty$, $A_{n-1,k+1}$ tends to $A_{n,k+1}$; so putting it all together, we arrive at

$$\left(1 + \frac{1}{k+1}\right) A_{n,k+1} \sim \frac{n^{k+2}}{k+1},$$

or,

$$A_{n,k+1} \sim \frac{n^{k+2}}{k+2}.$$

By restoring the powers of n in the definition of $I_{n,k}$, it is easy to rewrite this in the form $I_k = kI_{k-1}/(k+1)$, i.e., $I_k = 1/(k+1)$.

Once again, there are methodological insights to be drawn from this first-ever evaluation of integrals of arbitrary positive integral powers. To begin with, there is the reliance on an enlightened first guess as an almost automatic initial step and on its subsequent refining. But instead of leading to a simple iteration, the purpose it serves here is as the starting point of a procedure that ends in relating a quantity indexed by k, I_k, to I_{k-1}, in other words a genuinely inductive procedure. We have a hint that this is how the calculation of I_k was meant to go in the deliberate foregoing of the use of the known values of $I_{n,k}$ to evaluate I_k for k upto 3 (even the sum $A_{n,1}$ of the arithmetic series of natural numbers is not cited); the *yukti* is designed to cover all k as the passages cited earlier suggest. The hint is reinforced by the terminology. The integral I_1 is called *mūlasaṃkalitam*, the root from which the *saṃkalitam*s of all higher powers are generated recursively (except in one place where it is referred to as *kevalasaṃkalitam*, simple integral, to dstinguish it from mutiple integrals). Related to this is the painstakingly exhaustive exposition of the entire computation, letting the student/reader see at every step that the method is logically rigorous – there is no guesswork here.[39]

A final insight concerns the reordering (or resummation) identity which I write, generalising slightly its form as used in the study of $I_{n,k}$, as

$$\sum_{i=1}^{n-1}(n-i)f_i = \sum_{i=1}^{n-1}\sum_{j=1}^{i} f_j$$

[39]Compare Newton on the evaluation of the coefficients in the sine series: "Let it be noted here, by the way, that when you know 5 or 6 terms of those roots you will for the most part be able to prolong them at will by observing analogies" (*De Analysi per Aequationes Numero Terminorum Infinitas* in [18] Vol.II, p.206). In one of his letters to Oldenburg (for onward transmision to Leibniz), the *Epistola Posterior* [19] Vol.II, p.130, Newton writes about how he "first chanced upon" the general binomial series as an extension of the results of Wallis by means of "intercalation" and "interpolation". On the question of the nature and role of proof in Indian mathematics, the reader may find it instructive to read Srinivas's essay in [Sarma] alongside Roddam Narasimha [3] cited earlier.

for arbitrary (say real) coefficients f_i. (Its elementary proof is the same as in the special case $f_i = i^k$). One recognises it for what it is by passing to the limit, rewriting it as an identity for integrals

$$\int_0^x (x-y)f(y)dy = \int_0^x dy \int_0^y f(z)dz$$

of real functions f. To prove it the modern way, define

$$g(x) := \int_0^x f(y)dy$$

so that, by the fundamental theorem, $f(x) = dg(x)/dx$. This enables us to integrate the second term on the left by parts. The left side is then

$$xg(x) - \int_0^x y\frac{dg(y)}{dy}dy = xg(x) - xg(x) + \int_0^x g(y)dy = \int_0^x dy \int_0^y f(z)dz$$

which is the right side.[40]

3.5. Trial and Error

Having found an exact expression for $\pi/4$ as an infinite series of alternating reciprocals of odd integers, *YB* concludes chapter 6 with a section which Tampuran and Ayyar entitle accurately "Refinements needed to lighten (or shorten) the work of computing the circumference – their justification (*upapatti*) and accuracy".[41] These refinements take the form of estimating the remainder after a finite number of terms of the series by a trial-and-error procedure of *saṃskāram* leading to two different outcomes: alternative, faster-converging exact infinite series or approximate but extremely accurate estimates for the remainder as low degree rational functions of the termination point. True to its general style, *YB* does not say why it is absolutely

[40] The essential point of this exercise is that the finite version of the fundamental theorem is self-evident – the whole is the sum of its parts – and is never mentioned in any form in *YB*. This is an interesting aspect of doing as much work as possible before the limit is taken: it deemphasises the role of the fundamental theorem.

[41] In Sarma's translation, this section (6.8) is headed "*Antya-saṃskāra*: Final correction terms". The section opens with the sentence: "[I] describe [now] how the results obtained by dividing by higher and higher odd numbers are refined to make them come closer to the circumference by performing a final *saṃskāram*". The terminal clause can alternatively be read as ". . . by performing a *saṃskāram* after the last term". The expression *antya-saṃskāra*, with its funerary connotation ('the last rites') does not occur in this section nor, in my reading, anywhere else in the book. What is used here instead of *antya* (and elsewhere in the section to denote the term at which truncation is effected) is an equivalent non-Sanskrit Malayalam word having the same literal meaning ('final' or 'in the end') but not the ritual connotation.

essential to look for methods of reducing the computational burden, but this is one motivation that is easy to guess. The π series is unfit for *vyavahāram*, on account of its slow convergence – the first 56 terms add up [TA] to a value which is wrong in the third significant figure, much worse as an approximation than Āryabhaṭa's *āsanna* rational value which is correct to the fifth significant figure. For an endeavour that had its origin in dissatisfaction with rational approximations to π, this must have been galling.

In every other respect, however, this section is completely out of character. For the first and only time in the book, it has a good pedagogical discussion of matters that can be considered genuinely algebraic, namely operations on polynomials and rational functions. Since that discussion, fleeting as it is, is modelled on arithmetic with decimal numbers, it will be given its due briefly in section **5.1**. More remarkably, after explaining at length the 'philosophy' of an ingenious and novel semi-empirical technique of *saṃskāram*, *YB* contents itself with listing several results the technique is said to lead to by quoting (Sanskrit verses) from *TS*, without proof or any sort of justification – there is no *yukti* in these passages and little *bhāṣā*.

The conspicuous absence of *yukti* has encouraged some modern commentators to attempt to supply one, and some of the work in that direction has led to an intriguing situation. For instance, C. T. Rajagopal (one of the pioneer scholars of *YB*) and M. S. Rangachari have pointed out [26] that there is a beautifully recursive procedure to evaluate the truncation error *exactly* as an infinite continued fraction whose first few approximants coincide precisely with *YB*'s recipés. So, did Mādhavan or Nīlakaṇṭhan use such methods to arrive at their spectacularly effective results[42] which *YB* for reasons unknown decided not to elaborate? Within this mystery lies another riddle. Indian mathematics as a whole seems curiously unwilling to employ continued fraction expansions, a recursive device *par excellence*, especially as generated by the Euclidean algorithm. (Its rudimentary use in *kuṭṭaka* seems to be the exception that proves the rule). As of now, there is no satisfactory explanation; how these approximations were arrived at remains something of an enigma.[43]

The philosophy of the refining process is, briefly, as follows. *YB* first expresses the remainder after summing up to the ith term (with denominator $2i - 1 =: j$; so j runs through the positive odd integers) as $1/r_j$ with sign opposite to that of the last summand (r_j is therefore positive for all j). Then the remainder after truncation

[42] The third and final (as far as *YB* is concerned) approximant for the correction after 56 terms improves the accuracy of π from the second to the eleventh significant figure [TA].

[43] Rajagopal and Rangachari are, very properly, cautious on this question: they preface their continued fraction solution of the equation satisfied by the remainder with the sentence, "Let us make some observations which (in their main part) go beyond the evidence of the surviving Keralese texts but are perhaps not irrelevant" [26].

one term earlier is given by the recursive equation

$$\frac{1}{r_{j-2}} = \frac{1}{j} - \frac{1}{r_j}.$$

The basic idea is to think of the sum of the first $(j+1)/2$ terms of the π series for a chosen j as the first guess in a process of *saṃskāram* and the remainder $1/r_j$ as the correction resulting from the would-be *saṃskāram*. The equation above for the remainders (the error equation) is the condition for the *saṃskāram* to give the exact value for the remainder and the aim is to use it as a guide to arrive at a 'good' approximation. After recognising that an exact solution (as a finitely expressed function of j) is probably impossible, the discusion proceeds to make a patently absurd first guess, but now for r_j, namely $r_{j-2} = r_j = 2j$, and rejects it. The reason given, in a first explicitly functional piece of argument (the notion of a function in a geometric setting is implicit in the generalisation of the π series to the arctangent series, but here there is no longer any geometry to fall back upon), is that the remainder after the j-term must depend on j "in the same manner" as the remainder after the $(j-2)$-term depends on $j-2$, i.e., that r must be a (non-constant) function of j. So *YB* now offers a revised, functionally consistent guess, $r_j = 2(j+1)+x$, $r_{j-2} = 2(j-1)+x$, which has the property that $r_j - r_{j-2} = 4$, independently of j and x. This of course does not solve the error equation but only an approximate version of it,

$$\frac{1}{r_{j-2}} + \frac{1}{r_j} = \frac{1}{j - \frac{1}{j}}.$$

But it is a very good approximation for $j >> 1/j$ – for truncation after 56 terms for instance, j is of the order of 100 and the accuracy is roughly 1 part in 10,000. (*YB* actually defines and computes the inaccuracy as the difference $1/r_{j-2}+1/r_j-1/j$).

The search for even better approximations is now carried further. The variant $r_j = 2j+1, r_{j-2} = 2j-3$ (keeping $r_j - r_{j-2} = 4$; no reason for doing so is given) is tried and, once again, rejected since the inaccuracy is now one order in j higher ("in the second place in j"; the reference is to a place-value notation with variable base j for polynomial functions of j, see section **5.1**). At this point, the account gives up on the condition of constant difference between r_j and r_{j-2} and suggests solutions of the form $r_j = 2j+2+x/(2j+2)$, $r_{j-2} = 2j-2+x/(2j-2)$, settling finally on $x = 4$ as it keeps the error "in the first place in j", ("as said by the preceptor (*ācāryan*)", Nīlakaṇṭhan, Jyeṣṭhadevan's guru). The approximations for $\pi/4$ resulting from these prescriptions were first written out in [TA] and, in English, can be found in [RSS].

Hereabouts, *YB* introduces yet another recursive novelty and, simultaneously, chooses once again not to remain true to its title by citing several results from *TS* without any *yukti* accompanying them. The novelty consists in turning a given

saṃskāram for the remainders into a reordered infinite series for π by substituting the chosen correction terms for each of the terms in the original series, i.e., by replacing $1/j$ by $1/r_{j-2} + 1/r_j$, resulting in

$$\frac{\pi}{4} = 1 - \left(\frac{1}{r_1} + \frac{1}{r_3}\right) + \left(\frac{1}{r_3} + \frac{1}{r_5}\right) - \cdots .$$

If the r_j satisfy the error equation, this is an exact but logically trivial result: $\pi/4 = 1 - 1/r_1$; it is just the definition of r_1. It becomes interesting only if the r_j are approximate and *do not* satisfy the equation. But then it is no longer clear that the series sum up correctly to the value $\pi/4$, though one may suspect, correctly as we know now, that they will if the r_j are sufficiently fine (*sūkṣmam*) approximations. The series resulting from various choices for r_j (see [TA], [26] or [RSS]) are quite remarkable and some of them have been linked, *via* continued fractions, to certain results of Euler and of Ramanujan [26].

From the recursive perspective, this approach is an almost perfect illustration of the general philosophy of *saṃskāram*: one begins with an identity, namely the series for π in terms of r_j, then approximates each r_j in such a way that the errors are always of order higher than j and will (hopefully, since the procedure is very formal) cancel out in the end. Indeed, by the time chapter 6 is brought to a conclusion, we have been exposed to a whole sequence of *saṃskārams*, nested one within another, of varying degrees of *sūkṣmata*. The concluding passage is yet another verse from *TS*, containing yet another approximation, finer by one degree in j – the third approximant in the continued fraction expansion [26] of $1/r_j$ – leaving the question of where it all really came from tantalisingly open.

4. Varieties of Recursion – the Sine/Cosine Series

4.1. From Āryabhaṭa to Mādhavan

If the π series and the arctangent series are the logical culmination of attempts to improve upon Āryabhaṭa's approximate rational value for π, the sine and cosine series are, in the same measure, the final goal of the search that began with his sine table. Already in *AB*, the two problems are closely linked: without a sufficiently precise value for π, a table of sines accurate to one minute of arc could not have been made, no matter how finely the circle is divided. Indeed, the two verses setting out the general idea and the recipé for the computation of the sine differences (*Gaṇitapāda* 11 and 12) are immediately preceded by the π stanza. And, in going beyond Āryabhaṭa's level of precision, there would have been little point in Mādhavan's fine elaboration of a theory of trigonometric functions without a correspondingly fine and precise value for π. From these two interlinked perspectives therefore, the parallel development of the π series and the sine series in *YB* is very natural.

The surprise then is that the breakthrough took so long. The two Bhāskaras, both devoted Āryabhaṭa followers, seem to have missed the deeper meaning of *Gaṇitapāda* 11 instructing the listener (or reader) to divide the circumference so as to make it possible to find as many sines of equal arcs as one likes (*yatheṣṭa*). Whatever the reason for this oversight, and that is another story, it took until the advent of Mādhavan for Āryabhaṭa's vision to be finally realised.

The accurate/exact evaluation of π having been dealt with in chapter 6, the first half of chapter 7 of *YB* (upto and including section 7.8 of [Sarma]) is concerned wholly with the circle of ideas pertaining to the accurate/exact evaluation of $\sin\theta$ for an arbitrary value of θ. Quite remarkably, the first topic addressed is the original Āryabhaṭan numerical table of sine differences (of limited accuracy, necessarily) for values of $\theta = m\pi/48$, $m = 0, 1, \cdots, 24$, covering the first quadrant. (I shall use ϵ to denote the angle $\pi/48$, s_m for $\sin m\epsilon$, δs_m for the difference $s_m - s_{m-1}$ and $\delta^2 s_m$ for the second difference $\delta s_{m+1} - \delta s_m = s_{m+1} + s_{m-1} - 2s_m$). One might have thought that this was wasted labour for the author, knowing as he did that an infinitely accurate procedure which worked for any angle was soon going to follow. Apart from showing reverence for knowledge from earlier times ("*pūrvaśāstra*", meaning really *AB*), and familiarising the reader with the algebraic properties of differences and second differences of functions in a familiar situation, this may have had a practical motivation: the computation of the table using Āryabhaṭa's formula or the equivalent procedure given in *YB* takes much less effort than working with power series. In fact the section of *YB* containing the *yukti* behind the table is followed by the description of Mādhavan's interpolation formulae for sines and cosines of angles lying between two tabulated angles making it clear that, for purposes of astronomical calculations, there was no need to go any further and also that the sine table was an essential first step. We are once again made aware that the development of the power series owed less to practical needs than to the attraction of a problem that was there and that could be solved.

In any case, from the point of view of this article, the interest is in the fact that all three phases of that enquiry, tabulation, interpolation and power series expansion, are addressed by recursive methods of increasing subtlety.

4.2. Making the Sine Table

One point of departure for Āryabhaṭa's sine table (the expression will stand for a table of either the sine differences or the sines unless explicitly distinguished) is the value of $\sin\pi/6$; this is attested by (the second line of) *Gaṇita* 9, to the effect that the side of a regular hexagon is equal to the radius of its circumscribing circle. Thus the *Gaṇita* stanzas 9 to 12 take us along one logical sequence of steps leading to the table, beginning with $\sin\pi/6 = 1/2$, through the approximate value of π (3.1416) and the somewhat enigmatic (at first reading) verse 11 instructing us to divide the circumference into as many equal arcs as we please, and ending with the formula for

δs_m. In view of the content of verse 11, it is pertinent to wonder why this particular subdivision was chosen and why it stopped at $\pi/48$. As often suggested (first by Playfair [20] on the basis of his acquaintance with *Sūryasiddhānta*), it would have been straightforward to compute $\sin \pi/48$ by halving $\pi/6$ successively three times and using the theorem of the diagonal to relate the chord of an arc to the chord of half that arc, a relation almost certainly known to Āryabhaṭa; *YB* explains the geometry and gives the formula at the start of chapter 7, sections 7.2.1 and 7.2.2 of [Sarma]. Indeed *YB* concludes this part by stating that the sines of some of the 24 Āryabhaṭan angles can be obtained in this way while "some others" (effectively all the rest) can be found by starting with $\sin \pi/4$ and similarly subdividing. The whole table could thus have been made by direct computation using a formula for $\sin \theta/2$ that Āryabhaṭa knew. Given this fact, it is even more pertinent to wonder why his table as well as his recipé for making it deals only with sine differences, not the sines themselves. The germ of a plausible answer may lie in a second observation of Playfair's, that $\pi/48$ is the largest angle in this sequence for which the chord is equal to the arc to a precision of one minute.

A table of sines is a staple of Indian astronomical texts, right until the time of *YB* and beyond, as it was an essential aid to the computation of the true positions of heavenly bodies. In fact, by around the 9th century, the table acquired a canonical *kaṭapaya* form and, at least in Kerala, seems to have been just transcribed from text to text. From very early times, what is most often tabulated are the sines, not their differences (presumably since those are what is directly employed in planetary calculations) but the formula itself, when it is stated, is always for the sine differences. Already in *Sūryasiddhānta*, we find the table presented (in stanzas II.17 - 22 in the version translated and edited by Burgess [27]) directly for the sines while the formula (in II.16) is for their differences[44]:

$$s_m - s_{m-1} = s_1 - \frac{1}{s_1}(s_1 + \cdots + s_{m-1}).$$

In all essentials this is the same as the more ambiguously expressed formula of *AB* as interpreted by K.S. Shukla in his annotated translation [28]. Shukla in fact gives several equivalent rewritings of the formula, due to various traditional commentators as they came to grips, each in his own way, with *Gaṇita* 12. All of them employ Āryabhaṭan units, namely minutes, for both an angle θ (equivalently the arc subtending that angle) and its sine (equivalently the half-chord of arc 2θ of the standard circle of circumference 21600 minutes and radius $21600/2\pi = 3438$

---

[44] In his commentary on these stanzas Burgess expresses the curious view that the table was prepared not by using the formula but by direct subdivision, the formula being obtained empirically by a subsequent inspection of the table. Why would anyone first state a rule that was not going to be used *and then* use a different rule, unstated, to get the same results?

minutes).[45] In these units and to an accuracy of 1 minute, $\delta s_1 = \sin\epsilon = \epsilon = 225$, $\delta s_2 = \sin 2\epsilon - \sin\epsilon = 224$ and so $\delta^2 s_1 = -1$. Consequently, a multiplicative $\delta^2 s_1$ disappears from all formulae. All of the early commentators except one incorporate the resulting numerical simplification (a minor mischief which is easily set right by restoring the exact formula for $\delta^2 s_1$) as well as the approximation $\sin\epsilon = \epsilon$ in their versions of the sine-difference formula.

The exception is Nīlakaṇṭhan. Reading more into *Gaṇita* 12 than is literally there, he interprets the verse in a form free from conventions and approximations as the exact equation[46]

$$\delta s_{m+1} = \delta s_m - \frac{s_m}{s_1}(\delta s_1 - \delta s_2)$$

which, for later reference, I rewrite in the explicit second-difference form

$$\delta^2 s_m = \frac{s_m}{s_1}\delta^2 s_1 .$$

Nīlakaṇṭhan has himself given a geometric proof of his formula which can be found, in modern notation, in Shukla [28] along with an equivalent trigonometric formulation. I recall that proof here after highlighting what appears to have been, at least by the time of the Kerala school but possibly from much earlier, the 'conducting thread' leading to the consideration of differences and second differences of sines.

---

[45] A tribute is due here to one of the less celebrated innovations of Āryabhaṭa, the recourse to a natural, angular unit to measure chords in terms of the fixed angular 'size' of the circumference. This is the converse of the current practice of measuring angles in radians and defining the sine as the corresponding half-chord in a circle of unit radius and is strictly equivalent to it; in particular the use of modern sines etc. in the present work in place of the traditional Rsines etc. is in no respect misleading. It is in measuring chords in minutes that the value of π comes into play just as we need π to calculate the circumference of the 'unit circle'.

[46] It is not difficult to see in this exact generalisation of the original approximate rule a parallel with the other instance in which Nīlakaṇṭhan has read Āryabhaṭa's mind, when he takes the one word *āsanna* to mean that π is irrational. Surprisingly, neither of these instances of Nīlakaṇṭhan's insight finds a mention in *YB*, surprising in the first case because it is uncharacteristic of *YB* not to state a general rule when it is available and, in the other, in view of the thoroughness of *YB*'s treatment of the complex of issues relating to the π series. Besides, how can it be that the author of *YB*, a work composed to provide an exposition of the new mathematics "according to *TS*" as announced in its very first sentence, is unaware or negligent of such illuminating wisdom in the teaching of his own direct preceptor? A possible answer may lie in the fact that both these examples of mind-reading occur in *Āryabhaṭīyabhāṣya* which Nīlakaṇṭhan wrote late in life and in the possibility that this work was composed *after YB*. Such a resolution is of some chronological interest. K.V.Sarma's earliest date for *YB* is around 1530, by which time Nīlakaṇṭhan would have reached the ripe old age of 85 or so. Knowing that he wrote at least two other works after his commentary on Āryabhaṭa (Sarma's introduction to[1]), one of them (the *vyākhyā* of his own *Siddhāntadarpaṇa*) of major epistemological importance, it is not unreasonable to assign to it a date well before 1530. Correspondingly, if this work postdates *YB* as suggested here, 1530 will be more like a strong *upper limit* date for *YB*.

That thread starts with the realisation of the inadequacy of the rule of three (*trairāśikam*) when dealing with quantities whose mutual dependence is not linear, in the present instance the 'sides' of 'similar triangles' when one 'side' is an arc rather than a chord of the circle. Just before the description of the computation of the sine table (at the end of section 7.4.1 of [Sarma]), *YB* has a brief but enlightening passage on the error of relying on the rule of three to determine chords from arcs. After warning that it will lead to gross (*sthūlam*) results, it ends with an explanation and an exhortation:

> The reason for this [grossness is as follows]. The second arc (2ϵ in our notation) [is] twofold the initial arc, the third arc threefold. Thus the arcs. The second chord is not twofold the initial chord, the third chord is not threefold, and so on. The reason for this [is as follows]. The initial arc has no curvature [and is] almost equal to the chord since the *śaram* (the versine, $1 - \cos\epsilon$; see figure 7.2 of [RSS] for the terminology) is small. . . .So do not do *trairāśikam* with respect to the arc because the result will be gross.

In effect we are told that $\sin\theta$ is not a linear function of θ since the relation between the arc and the chord is not one of proportionality. In particular, s_m is not the mean of s_{m+1} and s_{m-1} and the difference δs_m is not constant. The grossness in the value of s_m that will result from doing *trairāśkam* is thus measured by the difference between $s_{m+1} + s_{m-1}$ and $2s_m$ or equivalently between δs_{m+1} and δs_m, in other words the second difference. The line of thinking here is evidently very similar to what we have already met in other varieties of *saṃskāram*: the gross first guess is that the sine can be linearly interpolated, i.e., that the difference δs_m is independent of m, and the next step is to evaluate the deviation from linearity as the second difference.

Nīlakaṇṭhan does this geometrically by considering two astutely chosen pairs of similar right triangles (Shukla's trigonometric proof uses the addition formula for $\sin(m+1)\epsilon$ and $\sin(m-1)\epsilon$) leading to

$$\delta^2 s_m = 2s_m(\cos\epsilon - 1).$$

Nothing is more natural now than to eliminate $\cos\epsilon - 1$ in favour of $\delta^2 s_1$ and to arrive at the formula written down earlier – Nīlakaṇṭhan's equation is thus exact. If the initial values s_1 and s_2 (and hence $\delta^2 s_1 = s_2 - 2s_1$) are also exactly known, the solution will give the exact values of all $\delta^2 s_m$, all δs_m and all s_m recursively. Such an exact solution will satisfy all the symmetries of the sine function if the computation is carried beyond the first quadrant: $s_{25}^2 = 1 - s_1^2$, etc. and be periodic: $s_{97} = s_1$, etc. Here then is the quintessential illustration of what I have called cyclic recursion in section **1.7**.

Thus Nīlakaṇṭhan. As for Āryabhaṭa, while he may well have been led to second differences by the failure of the rule of three, it appears doubtful that he arrived at his formula along this precise path. Apart from the evidence of the actual words of *Gaṇita* 12, the construction invoked by Nīlakaṇṭhan was probably unknown

to Āryabhaṭa, being (along with several variants) a characteristic of the Kerala approach to circle geometry.[47] Every piece of writing bearing on this question points to Āryabhaṭa taking recourse, from the beginning, to the approximation $s_1 = \epsilon$ along with the known formula for doubling or halving the angle to evaluate $\delta^2 s_1$ as

$$\delta^2 s_1 = -2s_1.2\sin^2\frac{\epsilon}{2} = -s_1\epsilon^2$$

and proceeding thence recursively. More generally, Nīlakaṇṭhan's equation will take the corresponding approximate form

$$\delta^2 s_m = -s_m\epsilon^2.$$

There is no need to underline the enormous significance of this equation. As ϵ is further subdivided into arbitrarily (*yatheṣṭam*) small pieces ("atomised"), the approximation will tend to exactness, the difference equation to the correct differential equation for the sine and the two initial values $\sin 0 = 0$ and $\sin\epsilon = \epsilon$ to the appropriate initial conditions for its solution. Solving the equation also transforms into a subtle iterative process, yet another productive variant of the method of *saṃskāram*. This is one of the ways in which the sine and the cosine series were finally nailed down by Mādhavan. *YB*'s own treatment of the table is meticulously adapted from the start to a general division of the quadrant and to the eventual goal of making each arc segment infinitesimally small as we will see in section **4.4**.

4.3. The Interpolation Formula – Simultaneous *Saṃskāram*

The description of the sine table is immediately followed in *YB* by a section outlining how to use the tabulated values to estimate $\sin\theta$ and $\cos\theta$ when θ lies "in the gap" between two consecutive tabulated angles. There are hints that this topic may have been considered a distraction: it is brief to the point, almost, of impatience – the whole passage is two paragraphs and less than one printed page ([Sarma], section 7.4.3) long. The operative paragraph sketches a geometrical construction in terms of the inevitable pair of similar right triangles and indicates (barely) what to do with them in order to obtain a second order interpolation formula. The reader will find the geometry in Fig. 7.6 of [Sarma] and an explanation of the derivation in [RSS].

---

[47] Shukla's use of the addition formula is of course unnecessary and may be misleading. While the sine difference formula follows directly from the addition formula, the reverse implication is less direct. In any case, Āryabhaṭa almost certainly was not in possession of the general addition formula. The distinction of its discovery also belongs to Mādhavan, going by the laudatory references to his *jīveparaspara-nyāyam* (literally, the principle of adjacent chords; more informatively, the principle of the half-chords of adjacent arcs) and its heavy use by his followers (including in *YB*).

For an incremental angle $\delta := \theta - m\epsilon < \epsilon$, the formulae are

$$\sin(m\epsilon + \delta) = \sin m\epsilon + \delta \cos m\epsilon - \frac{1}{2}\delta^2 \sin m\epsilon,$$

$$\cos(m\epsilon + \delta) = \cos m\epsilon - \delta \sin m\epsilon - \frac{1}{2}\delta^2 \cos m\epsilon$$

and they depend essentially on the Āryabhaṭan approximation of $\sin \delta/2$ by $\delta/2$. The section ends with " If this is not of sufficient accuracy, take the full chord of one quarter of the residual arc (i.e., $\delta/4$ in place of $\delta/2$) . . . if this is also not sufficient, take half of that . . . " followed by a reference to a verse in *TS*. The paragraph does include the word *saṃskāram* but that is about all.

In fact the reasoning behind the interpolation is interestingly recursive. The sentence cited above asks us to take the computed approximate value of $\sin(m\epsilon + \delta/2)$ as the new input and, proceeding as before, to calculate the new output $\sin\theta$ as $\sin((m\epsilon + \delta/2) + \delta/2)$; obviously the process can be continued indefinitely. But it is to be stressed that while there is an infinite *saṃskāram* at work here, based on a nested sequence of segments of arc of length $\delta/2^i$ in geometric progression, the segments are not taken to zero at any stage – there is nothing infinitesimal, no calculus. Perhaps this accounts for the inhabitual abruptness with which *YB* terminates the discussion of the interpolation. To follow this particular thread, therefore, we have to rely on sources other than *YB*.

It is useful to start by recognising that the only point about making the expansion around the Āryabhaṭan angles $\theta = m\epsilon$ is that their sines and cosines are supposed known from the table; the geometry itself works for any value of θ. Similarly, though δ is taken to be between $m\epsilon$ and $(m+1)\epsilon$, the geometry works for any angle δ. With this in mind we turn to Kim Plofker's careful analysis ([29, 30]) of the far more extensive account of this topic by Jyeṣṭhadevan's younger contemporary Śaṅkaran in his commentaries on *TS*. As Plofker has underlined, the formulae are established for both the sine and the cosine simultaneously by setting up a pair of coupled equations for the correction: the output of alternating steps in the refining of the correction, say the ith for $\sin(\theta + \delta)$, involves $\cos(\theta + \delta/2^{i-1})$ whose evaluation involves $\sin(\theta + \delta/2^i)$ and so on. The expressions resulting from iterating the step *ad infinitum* (Śaṅkaran does the first seven of them) is strongly reminiscent of the nested brackets occurring in the treatment of the geometric series by means of the elementary recursive identity but more spectacular ([30]):

$$\sin(\theta+\delta) = \sin\theta + \frac{\delta}{2}\left(2\cos\theta - \frac{\delta}{2}\left(\sin\theta + \frac{\delta}{2}\left(\frac{\cos\theta}{2} - \frac{\delta}{2}\left(\frac{\sin\theta}{4} + \cdots\right)\right)\right)\right),$$

$$\cos(\theta+\delta) = \cos\theta - \frac{\delta}{2}\left(2\sin\theta + \frac{\delta}{2}\left(\cos\theta - \frac{\delta}{2}\left(\frac{\sin\theta}{2} + \frac{\delta}{2}\left(\frac{\cos\theta}{4} - \cdots\right)\right)\right)\right).$$

Removing the brackets turns them into infinite series in powers of δ (equations 28 of [30], extended by repeated substitutions, in a slightly different notation), the numerical coefficients being, naturally, increasing inverse powers of 2.

To our eyes these are strange series. The coefficients of successive powers of δ involve successive derivatives of the sine and the cosine linearly as in the Taylor series and the first two terms are exactly the same as in the Taylor series; but Taylor series they are not. We can begin to guess the precise source of the error (see the next-but-one paragraph) but that the series are wrong is made absolutely clear by the fact that they specialise at $\theta = 0$ to (equations 29 of [30])

$$\sin\delta = \delta - \frac{\delta^3}{2^3} + \frac{1}{4}\frac{\delta^5}{2^5} - \cdots,$$

$$\cos\delta = 1 - 2\left(\frac{\delta^2}{2^2} - \frac{1}{4}\frac{\delta^4}{2^4} + \cdots\right).$$

which are together inconsistent with the fact that sine and cosine are each other's differentials up to a sign, something certainly known to Mādhavan and much used in *YB*. And who could have failed to notice that they are inconsistent also with Mādhavan's own sine/cosine series?

Not only is the space devoted in *YB* to the subject of interpolation dismissively short, there are also no statements in that short space to the effect that an exact expansion for $\sin(\theta + \delta)$ will result if the recursive construction is continued forever (no "atomisation", no division by *parārddham*), statements of the sort found quite freely in connection with the π series (sections 6.3 and 6.4 of [Sarma] cited earlier) and the sine/cosine series (e.g., sections 7.5.2, 7.5.4 and 7.5.5). Is one to conclude then that for Jyeṣṭhadevan, unlike perhaps for Śaṅkaran, the formulae were no more than a reliable numerical approximation, a means of extending the utility of the sine table, with no deeper significance? If the answer is yes, that will be yet another piece of evidence for *YB*'s infallible instinct for what was genuinely deep and original in Mādhavan's accomplishments, but there may be a more prosaic explanation. The recursive element in this small section, in particular the idea of simultaneous *saṃskāram*, comes into its own when used in conjunction with Mādhavan's infinitesimal philosophy to generate the sine/cosine series. In that context it certainly gets its due attention, in the sections of *YB* that follow.

The point to be underlined at the end is that doing a sequence of iterative corrections of the type seen above is not by itself calculus, any more than the geometric series obtained in the same way or the general binomial series is calculus, irrespective of whether the interpolation series coincide with the Taylor series or not and, even, whether they are correct or not. The essence of the Kerala approach to calculus was the same as it was to be in Europe and easily described in the language of its modern version: divide the range of the independent variable (the tangent to an arc in the arctangent series and the arc in the sine series, next subsection) into n equal

parts with n eventually to be made to tend to infinity, linearise the dependent variable (the arc and the half-chord respectively in the two problems) as a function of the independent variable at each point of division and add up the linearised quantities ensuring that the remainders so omitted tend to zero in the limit, thereby implicitly establishing/employing the fundamental theorem. The interpolation formulae are outside that framework. The 'independent variable' δ is subjected to an iterated, nested subdivision as in the case of the geometric series (hence the similarly nested final expression) and as a result the segments of δ do not tend to zero in the limit (they are not differentials); one only has an infinite sequence of segments whose limit is zero. There is a process of linearisation, but the 'error' involved in that process is not controllable in the limit. At the fundamental level, therefore, there is no conceptual linkage between the two. These shortcomings could be overcome, and were, by Mādhavan himself in the course of developing the sine/cosine series around $\theta = 0$, but then it becomes calculus. He could also have rewritten the series, but did not, as expansions around any other point, true Taylor series [16], by noting that the length of an arc did not depend on its initial point being the direction east, that it was invariant under translations of the circle.

It is in this light that the controversy that erupted a few years ago around the question of whether calculus was invented in India (the debate can be followed starting with [30] and the references therein) is to be viewed. The misinterpretation was to some extent inevitable: it owed its origin to a reading of the wrong material (the interpolation formulae) from the wrong sources (Śaṅkara Variyar's elaborations on *TS*) while the right source remained, untranslated and largely unstudied, hidden in a language inaccessibe except to a few. Thanks finally to the long overdue translation of *YB* by K.V.Sarma, aided in no small measure by the pioneering work of Tampuran and Ayyar to all three of whom this article is a tribute, that reason is no longer there.

4.4. From Differences to Differentials

The development of the sine and cosine series

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots$$

and

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots$$

borrows recursive elements both from the construction of the sine table (working with sine differences) and from the interpolation formula (the method of simultaneous *saṃskāram*). The critical input however is the novel concept already put to

use in the theory of the arctangent series, the infinitesimal geometry of Mādhavan. Though the conceptual and the technical go hand in hand, even more inseparably in this part of *YB* than elsewhere, my emphasis will again be on the recursive techniques employed; the originality of the calculus component of the arguments will figure only incidentally, mostly in this subsection and in section **4.7**. So the description here of the geometry is minimal (greater detail can be found in [RSS]), only to the extent required to make the working out of the steps comprehensible and to highlight the specific steps that bear witness to the birth of calculus.

Thus we begin by considering a circle of unit radius and an arc of it subtending an angle θ at the centre. The arc is divided into a number of equal segments and the objects of study are the sines and cosines of the angles defined by the points of division, geometrically the half-chords through these points and their distances from the centre, *as well as* the corresponding quantities at the midpoints of the segments. Nothing is lost and some extra effort in writing saved if we divide θ into an even number $2n$ of segments of length $\delta := \theta/2n$ from the start; so the quantities of interest are $s_i := \sin i\delta$ and $c_i := \cos i\delta$, $i = 1, 2, \cdots 2n$.

The reason for considering the midpoints of the arc segments simultaneously with their endpoints becomes apparent quickly. Applying *trairāśkam* to two suitable pairs of similar triangles, *YB* shows that the difference in the sines (cosines) of two angles differing by two segments is proportional to the cosine (sine) of the intervening angle:

$$\delta s_i := s_{i+1} - s_{i-1} = 2s_1 c_i,$$

$$\delta c_i := c_{i+1} - c_{i-1} = -2s_1 s_i.$$

These exact, elementary and beautiful relations are thus the finite difference analogues of the differentials of the sine and the cosine. They are also true in the more general form

$$\sin(\phi + \chi) - \sin(\phi - \chi) = 2 \sin \chi \cos \phi$$

for any ϕ and χ as a quick glance at the geometry makes clear. Though no different in method or conclusion from Nīlakaṇṭhan's treatment of the tabulated sines (with the arc segment fixed at $\pi/48$), their general validity for an arbitrary division is now acknowledged and underlined for the first time: "Here the arc segment has to be imagined to be as small as one wants ... [but] since one has to explain in a certain [definite] way, [I] have said [so far] that a quadrant has twentyfour chords" ([Sarma] 7.5.2). Other issues addressed by this simple construction/observation include the quantification of the deviation from linearity (the inapplicability of the rule of three) and the fact that while the difference formulae follow from the addition formulae, the former can be established more directly and simply.

The second differences are defined in the same way: $\delta^2 s_i := \delta s_{i+1} - \delta s_{i-1}$ and similarly for the cosines. Because the sine and cosine are each other's first

differences, the second differences can be calculated without any further geometry:

$$\delta^2 s_i = 2s_1 \delta c_i = -4s_1^2 s_i,$$

$$\delta^2 c_i = -2s_1 \delta s_i = -4s_1^2 c_i.$$

Consequently the content of the second difference equation is identical to that of the pair of coupled first difference equation. *YB* deals with both sets, often in conjunction, in its search for the solution. The slight lack of focus that results – the treatment on the whole is less streamlined than that of the arctangent series – leaves one with the impression that this first record of the detailed understanding of what was after all a revolutionary advance was not completed without struggle. In any case, the critical fact that these equations (especially the second difference equation) which were originally derived for $2\delta = \pi/48$ (and for that value of δ are the same as Nīlakanṭhan's starting point for the sine table) remain valid for any value of δ gets its due attention as attested in the quotation above.

At this point it is convenient to break with the order in which *YB* broaches the questions that arise in finding the exact solutions in the limit and note the obvious fact that the equations themselves are well-defined in the limit as the familiar differential equations satisfied by the sine and the cosine. As has been said several times already, *YB*'s way is to first solve them recursively and then to take the limit of the solution. That will be described in the following two subsections. For the present, let us just note that in the limit of large n, the angular segment $\delta = i\theta/2n$ tends to 0, the chord length $s_1 = 2\sin\delta$ tends to 2δ, $i\delta$ becomes some general angle $\phi < \theta$ and the difference equations become the differential equations

$$\frac{ds(\phi)}{d\phi} = c(\phi),$$

$$\frac{dc(\phi)}{d\phi} = -s(\phi)$$

and

$$\frac{d^2 f(\phi)}{d\phi^2} = -f(\phi)$$

respectively for $f = s$ or c, defining $df(\phi)/d\phi := \lim_{\delta\to 0}(f(\phi+\delta) - f(\phi-\delta))/2\delta$. These equations cannot be solved by simply integrating their right sides unlike in the case of the arctangent series where the problem was reduced geometrically to a quadrature, namely solving the first order differential equation $d\theta(t)/dt = 1/(1+t^2)$, and then further reduced to evaluating integrals of powers by means of the recursively generated geometric series for the inhomogeneous term. The solution of the corresponding difference equations will face the same problem.

4.5. Recursive 'Integration' of a Difference Equation

To repeat, the difference equations for the sine and the cosine are homogeneous equations, involving the same unknown quantities on the right as are sought and so no direct evaluation or quadrature is possible. It is the strategy devised to get around this difficulty that makes the final result, the sine/cosine series, the most impressive achievement of *YB*'s calculus. Apart from the calculus itself, the strategy involves several recursive novelties of which the most fundamental is a process of *saṃskāram* which kicks off with the first approximation of replacing the unknown chords by the corresponding (known) arcs (i.e., $\sin\phi$ by ϕ) and repeatedly refining it.

As a preliminary to the *saṃskāram*, *YB* makes the elementary observation that, in general terms, for a function $f(x)$ and a division $(x_0, x_1, \cdots x_n)$ of the domain of x,

$$f(x_i) - f(x_0) = \delta f(x_i) + \delta f(x_{i-1}) + \cdots + \delta f(x_1),$$

with $\delta f(x_i) := f(x_i) - f(x_{i-1})$: the whole is the sum of its parts (the *twitsoip* principle? the fundamental theorem of pre-calculus?). For the sine function, this allows us to write

$$s_{2n} = \delta s_{2n-1} + \delta s_{2n-3} + \cdots + \delta s_1,$$

taking account of our symmetric definition of δ and the initial value $s_0 = 0$. The difference equation for s can now be used to rewrite this as

$$s_{2n} = 2s_1(c_{2n-1} + c_{2n-3} + \cdots + c_1) = 2s_1(c_{2n-1} + \cdots + c_3) + s_2.$$

But each c_i is also the sum of its parts,

$$c_{2i-1} = \delta c_{2i-2} + \cdots + \delta c_2 + c_1$$

and each δc_i can be substituted by using the difference equation for c, resulting in

$$s_{2n} = -4s_1^2(s_{2n-2} + 2s_{2n-4} + \cdots + (n-1)s_2) + ns_2.$$

An alternative way of writing this, useful later, is

$$s_{2n} = -4s_1^2((s_{2n-2} + \cdots + s_2) + (s_{2n-4} + \cdots + s_2) + \cdots + s_2) + ns_2.$$

The decomposition on the right into a series of series is designed to aid the use of a systematic process of *saṃskāram* and their sum is a measure of the deviation of the chord of the whole arc from the sum of the chords of the (eventually infinitesimal) arc segements. The corresponding equation for c can be derived by the same procedure. An alternative method is to start from the second difference equation and apply the

same *twitsoip* principle twice, to the second and first differences successively, to get to the same representation as above of s_{2n} (and c_{2n-1}). It seems appropriate to call it the *saṃkalitam* (a word used to mean both the finite sum and its limit as an integral in *YB*) representation.

What does this representation represent? Nothing more or less than the finite-difference counterpart of the formal integration of the relevant differential equation. For instance, in the limit $\delta \to 0$ of the representation

$$s_{2n} = 2s_1 \sum_{j=1}^{n} c_{2j-1},$$

$2s_1$ tends to $d\theta$ and the sum becomes the integral:

$$s(\theta) = \int^{\theta} c(\phi)d\phi.$$

Similarly the representation of s_{2n} as a sum of s_{2i} becomes

$$s(\theta) = -\int^{\theta} d\phi \int^{\phi} d\phi' s(\phi').$$

In other words, modulo the imposition of initial conditions, what is in play here is what becomes in the limit the technique of turning differential equations into integral equations by an appeal, usually tacit, to the fundamental theorem.

The *saṃkalitam* representation continues to hold when n is replaced by any $i < n$, since θ is an arbitrary angle. For all $i > 1$ it is a nontrivial expression of the content of the difference equations (for $i = 1$ it is empty: $s_2 = s_2$). But it is an auxiliary result – it does not solve the problem of finding $\sin\theta$ as an explicit function of θ. *YB* says: "Here [we have seen that] the difference between the arc and the chord is caused (or expressed) by all the chords that lie below the chosen arc (*iṣṭacāpam*). As [we] do not know any of these chords, think of the chord as the arc itself and do the *saṃkalitam* of the arc" ([Sarma] 7.5.2). So the first guess is to replace s_{2k} by $s_{2k}^{(1)} := 2\delta k$ for all $k < i$ and to compute s_{2i} (in particular s_{2n}) from the *saṃkalitam* representation. But before describing the result and its further refining, it seems best to define and fix the notation, once and for all, for certain combinatorial coefficients in terms of which they are expressed. These coefficients are recursively defined as

$$S_k(i) := \sum_{j=1}^{i} S_{k-1}(j)$$

with

$$S_1(i) := \sum_{j=1}^{i} j.$$

In particular $S_k(1) = 1$ for all k. They have simple algebraic expressions which have an interesting history and to which we will pay some attention soon but, for the present, it is enough to work with the definitions.

The *saṃskāram* is carried out as follows. In the first step set s_{2i} as $s_{2i}^{(1)} = 2i\delta$ for all $i < 1$ but treat the factor $4s_1^2 =: \alpha$ as a known parameter not subject to correction since there is no equation to determine it. With this input, the first approximation for s_{2n} is a sum-of-sums:

$$\begin{aligned} s_{2n}^{(1)} &= 2n\delta - 2\alpha\delta[((n-1)+\cdots+1)+((n-2)+\cdots+1)+\cdots+1] \\ &= 2n\delta - 2\alpha\delta(S_1(n-1)+\cdots+S_1(1)), \end{aligned}$$

and hence simplifies to

$$s_{2n}^{(1)} = 2\delta n - 2\delta\alpha S_2(n-1)).$$

In the next step, the input values $s_{2i}^{(2)}$ for $i = 2, \cdots, n-1$ (s_2 does not change since there is no equation to determine it) are chosen as the output of updating the preceding sines:

$$s_{2i}^{(2)} = 2\delta(i - \alpha S_2(i-1)).$$

Hence the second approximation for s_{2n} is

$$\begin{aligned} s_{2n}^{(2)} \;=\; & 2n\delta - 2\alpha\delta\{[((n-1) - \alpha S_2(n-2)) + \cdots + (1 - \alpha S_2(0))] + [((n-2) \\ & - \alpha S_2(n-3)) + \cdots + (1 - \alpha S_2(0))] + \cdots + [(1 - \alpha S_2(0))]\}. \end{aligned}$$

Using successively the recursive definitions of S_3 and S_4, we get the final form of the second *saṃskāram* as

$$s_{2n}^{(2)} = 2n\delta - 2\alpha\delta S_2(n-1) + 2\alpha^2\delta S_4(n-2).$$

The *saṃkalitam* representation being linear, it is straightforward to extend the process indefinitely (no need to invoke analogies). The final result is an expansion of s_{2n} in powers of α:

$$s_{2n} = 2\delta n - 2\delta\alpha S_2(n-1) + \cdots + 2\delta(-1)^k\alpha^k S_{2k}(n-k) + \cdots.$$

This is quite a remarkable formula. For any finite n it terminates at $k = n-1$ since $S_k(0) = 0$ by definition for any k but that will not give the exact value of

s_{2n} because the error in the first guess of replacing chords by arcs will propagate through the iterations. One can undo the error by choosing $s_{2i} = is_2$ (i.e., i times the first chord segment rather than the arc) as the first guess and carry out further refinings as before, resulting in

$$\sin\theta = \sin\frac{\theta}{n}\left(n - 4\sin^2\frac{\theta}{2n}S_2(n-1) + 4^2\sin^4\frac{\theta}{2n}S_4(n-2) + \cdots\right).$$

For $n = 1$ we get the tautology $\sin\theta = \sin\theta$ as expected, for $n = 2$ the Āryabhaṭan relation $\sin\theta = 2\sin(\theta/2)(1 - 2\sin^2\theta/4)$, etc. In fact the formula is nothing but the generalisation of Nīlakaṇṭhan's exact solution for the sine table to arbitrary divisions of an arbitrary angle.

YB does not quite get to describing the finite series. It has, once again, more ambitious goals than a clever trigonometric identity. After setting out the *saṃskāram*, it straightaway takes the limit of large n (for fixed θ) in all the trigonometric factors occurring in the series (with the usual reference to *parārddham* and "endless atomisation"): $2\delta \to \theta/n, \alpha \to \theta^2/n^2$, leaving the coefficients $S_{2k}(n-k)$ untouched for the moment. As n tends to infinity, the finite series turns into a power series whose dominant terms give

$$\sin\theta \sim \theta - \frac{\theta^3}{n^3}S_2(n-1) + \frac{\theta^5}{n^5}S_4(n-2) - \cdots.$$

The asymptotic form of $S_k(n)$ is actually easy to find and already available from chapter 6 of *YB*: $S_k(n) \sim n^{k+1}/(k+1)!$, leading finally to the sine series.

The derivation of the cosine series proceeds in parallel. Also, instead of solving the second-difference equation for the sine directly, we could have started from the two first-difference equations for the sine and the cosine, written down the *saṃkalitam* representations for them, made the first guesses $\sin\theta = \theta$ and $\cos\theta = 1$, refined them in alternation and arrived at the same final series.

The *saṃskāram* technique can be used even more readily *after* taking the limit, to solve the resulting differential equations. As noted earlier, we can appeal to the fundamental theorem to formally integrate the differential equations to the integral form of the *saṃkalitam* representations; for instance the second order equation for the sine becomes the integral equation

$$\sin\theta = \theta - \int_0^{\theta} d\phi_1 \int_0^{\phi_1} d\phi_2 \sin\phi_2,$$

on supplying the limits and constants of integration and using the initial conditions $\sin 0 = 0$, $\cos 0 = 1$. The *saṃskāram* way of solving it is to substitute the first guess $\sin\phi_1 = \phi_1$ in the integrand on the right to produce $\sin\theta = \theta - \theta^3/3!$, then substitute this expression back into the integrand and so on. Alternatively and even more faithfully to the three *saṃskāram* steps listed in section **3.1**, replace the sine

in the integrand by its integral representation repeatedly. The result after k steps is the identity

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \cdots + (-1)^k \int_0^\theta d\phi_1 \int_0^{\phi_1} d\phi_2 \cdots \int_0^{\phi_{2k-1}} d\phi_{2k} \sin\phi_{2k},$$

which, in structure, resembles the elementary recursive identity first encountered in section **3.3** but by now has evolved into a more intricate *saṃskāram* equation. We know that the last (error) term vanishes asymptotically but, unlike when dealing with the arctangent series, *YB* does not talk about the convergence of the sine/cosine series.

4.6. An Inductive Proof That Is Absent

In some contrast to its treatment of the coefficients in the arctangent series where their inductive evaluation was done after their asymptotically dominant form was determined, *YB* describes exact formulae for the coefficients $S_k(n)$ for general k and n. It starts by giving the traditional geometric *yukti* for sums of natural numbers $S_1(n) = n(n+1)/2$ in surprising detail and then goes on to state the rule for $S_2(n)$ and $S_k(n)$ more generally: "Multiply together the *padam* (the whole arc, in my notation n), *padam* plus one and *padam* plus two. The result of dividing it by the product of one, two and three is the second *saṃkalitam*. In the same manner multiply the numbers which are increased by one at a time and divide it by the product of the same number [of factors] one, two, etc. The result is the *saṃkalitam* which is one below" ([Sarma] 7.5.3):

$$S_k(n) = \frac{n(n+1)\cdots(n+k)}{(k+1)!}.$$

Why does *YB* take the trouble of stating the exact formula for $S_k(n)$ when all it is interested in is its asymptotic form $S_k(n) \sim n^{k+1}/(k+1)!$ which was already obtained and stated in chapter 6? The reason, almost certainly, was the fact that it was known. According to [RSS], the general formula appears in a work of Nārāyaṇa of the mid-14th century, which probably makes it its earliest occurrence. Slightly later than *YB*, the formula is also set out in Śaṅkaran's *Kriyākramakarī*, a commentary not on *TS* but on Bhāskara's *Līlāvati*. But none of these sources contains a *yukti* for the formula and that includes *YB*. This is especially surprising in the case of *YB* as an inductive proof, very much in line with what it has already accomplished and not significantly more challenging, can be readily constructed.[48] A modern formulation of the proof will go as follows.

[48] To my knowledge, none of the commentators on *YB* have supplied a proof either. [TA] verifies the formula for k up to 3, by resorting to the known sums of low powers of i which obviously could

Define

$$T_k(n) := \frac{n(n+1)\cdots(n+k)}{(k+1)!}.$$

Then

$$\begin{aligned} T_k(n) - T_k(n-1) &= \frac{n(n+1)\cdots(n-1+k)}{(k+1)!}((n+k)-(n-1)) \\ &= \frac{n(n+1)\cdots(n+k-1)}{k!} = T_{k-1}(n). \end{aligned}$$

But from the recursive definition of $S_k(n)$ we have directly $S_k(n) - S_k(n-1) = S_{k-1}(n)$. So the result $S_k(n) = T_k(n)$ will follow for given k and n if $S_{k-1}(n) = T_{k-1}(n)$ and $S_k(n-1) = T_k(n-1)$. It is sufficient therefore to show that $S_k(1) = T_k(1)$ for all k and $S_1(n) = T_1(n)$ for all n. Now i) $S_k(1)$ consists of the single term $S_{k-1}(1)$ and so on: $S_k(1) = \cdots = S_1(1) = 1 = T_k(1)$; and ii) (by definition), $T_n(1) = n(n+1)/2 = S_n(1)$ (as *YB* has taken the trouble to show geometrically). End of proof.

Even if writing out this proof serves no other purpose, it will have convinced the reader that it certainly was not beyond the mathematical (and literary) powers of *YB*: it is quite as easy to turn it around and present it in *YB*'s preferred style, starting with $k = n = 1$ and building it up one step at a time in k and n.

In seeking to identify the reason(s) for the general silence on the *yukti*, we may discount the notion that there was in fact no *yukti* and that the general formula was an extrapolation from small values of $k = 1, 2$ (known since Āryabhaṭa, *Gaṇita* 21 of *AB*) – the Kerala mathematicians most particularly were not in the habit of stating results lacking sound logical support. Śaṅkaran's enunciation of the general formula for $S_k(n)$, more precise than *YB*'s, is actually followed by the line "The *yukti* of this is not easy to follow (not *sugama*) and so is not described here". (The whole verse will be found in [6] p.248). And it is preceded by a geometrical demonstration of the cases $k = 1, 2$ involving a clever cutting up of the area and volume of suitable two (the same as in *YB*) and three dimensional figures. As Sarasvati Amma [6] wonders, how can such a demonstration be extended to $k > 2$ "unless one were to conceive spaces with more than three dimensions", especially in a work professedly meant for "the benefit of the not so intelligent" (*alpadhiyām hita*)?

It is tempting to suggest that what the slow-witted really found difficult was the transition from a long-familiar geometrical and arithmetical mode of reasoning to one based on combinatorial algebra and, even more disorientingly, on novel logical uses of recursive thinking that an inductive proof represented. That would

---

not be generalised since sums of general powers were not known. I got the inductive proof below from Chandrashekhar Cowsik to whom I express my thanks.

presuppose that such proofs were a genuine breakthrough, born of a conscious recognition that geometric methods had hit their natural limitations and that new *yukti*-techniques were called for. The extreme care with which the first recorded (as far as we know now) inductive *yukti*, for the terms in the π series, is constructed and presented in *YB* is then fully justified – the sums of powers greater than two that occur there are as resistant to geometry as the coefficients $S_k(n)$. And if the only *yukti* available for the computation of the latter was in fact inductive, we will have to conclude that Nārāyaṇa who lived just before Mādhavan, we are not certain where, was an adept of such techniques.

All this only deepens the mystery of the immediate antecedents of Mādhavan's achievements, he himself having left too little of his own to throw any light on it. We are all the more fortunate then that so many of his followers turned out to be such prolific authors.

5. Two Footnotes

5.1. Decimal Numbers and Polynomials

One of the most interesting sections of *YB* not directly tied to calculus is concerned with the development of a notation and a proper set of rules for dealing with algebraic operations on functions involving powers of one variable. The context in which the need for such a general algebraic machinery emerges is that of the estimation of the corrections to be applied when the π series is truncated. As described in section **3.5**, the remainder r_j after $i = (j + 1)/2$ (j a positive odd integer) terms satisfies a recursive equation in j and r_j is to be chosen, effectively by trial and error, so as to satisfy this equation as closely as possible. It is in this connection that *YB* makes its only explicit allusion to the notion of a function, also as noted in section **3.5**.[49] The truncation point not being fixed in advance, j is treated as an *avyaktarāśi*, "the unknown", taking values in positive odd integers. *YB* says ([Sarma] 6.8, p.74):

> Then [one] has to know (or learn) how to do computations (the expression used applies to a traditional method of calculating with tokens placed in compartments drawn on the floor, on a piece of cloth, etc.) with quantities without a [fixed] numerical value. When [such an] unknown is assigned a numerical value, a number [which was] in a [given] place occupies the next higher place [when multiplied] by that number. Thus are the place differences to be thought of, not tenfold as by one, ten, hundred, etc.$\cdots$.

[49] This is the closest *YB* comes to working from a genuinely abstract algebraic standpoint, and it marks a clear advance over Bhāskara's *Bījagaṇita*. Those who will wish one day to enquire into the direction Indian mathematics might have taken after Kerala, had historical circumstances been different, will find this part of *YB* a fascinating piece of text to parse and to ponder over.

So a natural number n in the first place stands for n itself, in the second place for $n \times j$ and so on. When the coefficients are fractions, they can be reduced to a common denominator, i.e., an overall numerical divisor, and for negative coefficients, *YB* cites Brahmagupta's rules for multiplication among numbers "of a positive or a negative nature" (*dhanabhūtam* or *ṛṇabhūtam*). What results is a notation that characterises a polynomial function of one variable (over the rationals) by the numbers in the compartments, i.e., by the ordered sequence of its coefficients which, up to a common denominator, can be a positive or negative integer – thus accounting for *YB*'s sudden interest in Brahmagupta's old rule for multiplication with negatives.[50]

The functions that turn up in the *saṃskāram* work are not polynomials but involve (small) powers of j and $1/j$. The error after truncation at j:

$$e(j) := \frac{1}{r_{j-2}} + \frac{1}{r_j} - \frac{1}{j}$$

for the trial choice $r_{j-2} = 2j - 2, r_j = 2j + 2$ is an example. Such functions are also to be reduced to a common denominator by extending the usual numerical method, which involves addition and multiplication, to polynomials. The result will be a rational function of which the numerator and denominator are treated independently. The algebraic needs of the operations on the functions occurring in all the variants of the correction techniques met with in section **3.5** are thus fully met if the operations of addition and multiplication are defined for polynomials. *YB* proceeds to explain how that is to be done and understood with the help of relevant examples. I quote the first part of that passage ([Sarma] 6.8, p.75).

> Here [I] show how the work is to be done when the unknown (*rāśi*, the reader will have noted that this word is used in two different senses in this section) is taken to be the last odd number (the truncation point j). Make (literally, "write" or "draw") two rows of compartments (*khaṇḍam*) so that each compartment encloses one place. In this the upper row is for the numerators and the lower row for the denominators. The odd number (the unknown j) is |1|0| (I have used vertical bars to separate the compartments). If the *rāśi* is negative (since j is always positive, what is meant here are the coefficients of powers of j) it must be distinguished by some mark. For zero, whatever thing is used for it.
>
> The divisor in the first correction term ($1/(2j-2)$) is two less than twice the unknown. For it, two in the second place and negative two in the first place: |2|2 ∗ | (where ∗ marks the negatives). Then the divisor in the second correction term is two more than twice the unknown. For that, in the second place, twice the *rāśi*, that is two and, in the first place, positive two: |2|2|.

[50] It is ironic that on the unique occasion on which *YB* resorts to a symbolic representation of a mathematical object, the resulting notation is more compact than the modern way of writing out a polynomial.

The passage continues with a description of the multiplication and division of a polynomial by a number (obvious) and the multiplication of two polynomials (relying on the fact that j is actually a number though unknown). There are no surprises here; the time and care spent on the explanation are perhaps, once again, warranted by the conceptual innovation.

At this early stage of our appreciation of the riches to be mined from a detailed exploration of *YB*, there is nothing much to add except perhaps to stress again how closely the notation for and the algebraic operations with polynomials are modelled on the the writing of numbers in the decimal place value notation and arithmetical operations with them. A polynomial is uniquely characterised by the coefficients of powers of the unknown. The unknown (variable) is left implicit just as the base 10 is in the decimal notation; it is not assigned a symbol but has a name, (*avyakta*)*rāśi*. The only difference is that the coefficients are from the ring of integers and negative coefficients have to be explicitly distinguished whereas in the decimal representation of numbers, each coefficient is a nonnegative integer modulo 10 – there are no carry over rules in the algebra of polynomials. We may nevertheless feel confident in the guess [16] that, for the Kerala mathematicians, the decimal numbers served as a bridge not only to the old idea of the infinite, but also to the acceptance of infinite series both numerical and functional, just as they did for Newton (section **1.4**).

5.2. The Roots of *Saṃskāram*, Roots by *Saṃskāram*

How far back in antiquity can we trace recursive computational techniques of the sort we have come across in *YB*, more particularly in their *saṃskaram* form? There are very credible indications that the idea of making an initial incorrect guess at the solution of a problem which is then sequentially refined to any desired degree of accuracy is very old indeed, as old as the earliest recorded mathematics in India, the *Śulbasūtra* manuals (8th-7th C. BCE) [31]. Three of them, those of Baudhāyana, Āpastamba and Kātyāyana, have an identically expressed rational numerical approximation for the square root of 2 which is most easily and logically obtained (the specific arithmetical formula given in the texts is a good guide in this) by an exercise in the method of *saṃskāram*. The technique also seems to be at the back of a rational algebraic formula for the approximate square root of a general nonsquare positive integer n found in the much later Bakhshali manuscript. It seems best to discuss these together, but in anti-chronological order, beginning with the general case.

The period of the Bakhshali manuscript has been a contentious issue, various scholars suggesting various dates ranging between the 1st and the 13th centuries CE. According to Takao Hayashi whose critical edition [15] contains the most thorough study we have of the work, it was most likely composed in the 8th century. (That makes it substantially more than a thousand years younger than the *Śulba* manuals).

Our interest is in one *sūtra* (*sūtra* Q2 in Hayashi's nomenclature) which is cited on more than one occasion in the surviving fragment of the text and which contains the formula mentioned above. The literal meaning of the verse is not unambiguous but there is unanimity on its mathematical content (see Hayashi [15]):

$$x := \sqrt{n} = m + \frac{r}{2m} - \frac{1}{2}\left(\frac{r}{2m}\right)^2 \frac{1}{m + \frac{r}{2m}},$$

where m^2 is the square number nearest to and less than n and r is the difference $n - m^2$. Bakhshali of course gives no proof or explanation but it is difficult to look at this formula without immediately recognising its *saṃskāram* credentials. In that spirit, the obvious first guess is $x_1 := \sqrt{n_1} = m$ so that r is a measure of the error. It is easy to reexpress n as the difference of two squares:

$$n = \left(m + \frac{r}{2m}\right)^2 - \left(\frac{r}{2m}\right)^2,$$

a type of reordering very popular in Indian mathematics. The next approximation x_2 is obtained by neglecting the second (and second order) term in the above exact formula:

$$x_2 := \sqrt{n_2} = m + \frac{r}{2m}.$$

Now define r' to be the difference between the exact value and the approximant x_2:

$$(m^2 + r)^{\frac{1}{2}} = m + \frac{r}{2m} + r'$$

which equation, on squaring and neglecting r'^2 but not $(r/2m)^2$, leads to

$$r' = -\left(\frac{r}{2m}\right)^2 \frac{1}{2\left(m + \frac{r}{2m}\right)}$$

and hence to the Bakhshali formula as x_3 in this sequence of approximations. The question of course is, is this how the author of *sūtra* Q2 arrived at the formula? . Since it is not obviously a result that should be "self-evident to the intelligent" (Nīlakaṇṭhan, alluding to the theorem of the diagonal), there presumably was a *yukti* and going by the choice of the 'variables' (the nearest perfect square and its deviaton from the required answer), we can at least tentatively assume that it was a recursive one, not very different from the one above. Indeed just such a justification has been suggested by commentators, the closest to the method of *saṃskāram* (though without drawing the parallel) being that of Channabasappa [32].

The iteration can be carried forward indefinitely by consistently neglecting the quadratic term in the error at each stage, again in the general spirit of recursive

refining. The result is easily expressible in an explicit infinitely recursive formula – a particularly neat and compact formulation is given in Hayashi [15], formula (11.1).[51] What is interesting now is that we can also evaluate $\sqrt{n}$ to arbitrary accuracy by expanding $\sqrt{n} = ((m + r/(2m))^2 - (r/(2m))^2)^{\frac{1}{2}}$ in powers of the second term in a (convergent) binomial series. The two series are not the same – successive terms have, in general, different structures for the denominators – and there is no reason why they should be; they are expansions in different variables. Nevertheless, the first three terms of the binomial series coincide exactly with the three terms in the Bakhshali formula. The first moral to be drawn from this is that the absence of a general binomial theorem was not much of a handicap in finding alternative (convergent in this case) infinite expansions, as noted earlier (the end of section **3.3**). The second comes from the similarity and dissimilarity with Mādhavan's interpolation formula (section **4.3**) which also agrees with the Taylor series in the first two terms: whether a given formal iterative expansion is a valid one is to be independently checked in each problem.

YB does not refer to the Bakhshali formula; it has no occasion to. But given the uncanny similarity of the only *yukti* we know for it to the general philosophy of *saṃskāram*, we are obliged to make doubly sure that this affinity is not something we have imposed on the actual, bare-bones, formula by overinterpreting it. In other words, how can we be sure that it was thought of as the starting point of an exact, endlessly recursive, procedure as the paragraphs above (and Hayashi's formula (11.1)) suggest, and not just as a clever and empirically reliable approximation without a general theoretical basis? It is here that we can turn for help to the numerical *Śulbasūtra* formula for $\sqrt{2}$. That formula is (literally reading the verses for the various numbers involved)

$$\sqrt{2} = 1 + \frac{1}{3} + \frac{1}{3 \times 4} - \frac{1}{3 \times 4 \times 34}.$$

On the other hand we can apply the Bakhshali formula to $\sqrt{2}$ ($m = 1, r = 1$) to get

$$\sqrt{2} = 1 + \frac{1}{2} - \frac{1}{12}$$

which is *numerically* equal to the sum of the first three terms of the *Śulbasūtra* formula. Even more strikingly, the next *saṃskāram* in our conjectured general *yukti* for Bakhshali leads to a correction term which is $-1/408$, again identical with the 4th *Śulbasūtra* term (discounting the particular factorisation employed).

The book of Sen and Bag describes several suggested reconstructions of the *Śulbasūtra* formula for $\sqrt{2}$ without mentioning the general Bakhshali formula. None of them agrees in every step with our conjectured Bakhshali *yukti*. But one

[51] Just to avoid any misunderstanding, such a general formula does not occur anywhere in Bakhshali.

due to L. Rodet, published in 1879, two years before the Bakhshali manuscript was even unearthed, comes very close. It starts with the first guess $\sqrt{2} = 1 + 1/3$ (rather than just 1) and in two steps of numerical *saṃskāram* gets to the *Śulbasūtra* approximation. It is obvious that Rodet chose the first two terms to reproduce the 1/3 term and it is equally obvious that the *dharma* of *saṃskāram* imposes no conditions on the first guess except those of practical utility (economy) and theoretical soundness (convergence, for instance). In any case, despite the numerical coincidence seen above, there is no way to be absolutely sure that the putative *yukti*(s) of our two sources were the same. The numbers which appear as the various terms and factors may or may not point to a particular method of arriving at the result; it is not impossible for example (though it appears unlikely) that the choice of the denominators $3, 3 \times 4$ and $3 \times 4 \times 34$, identically expressed in all three *Śulba* texts dealing with the question, was dictated by the demands of metrical composition.

It is also something of a surprise that it is the earlier text that carries the refining one step further than the later. Here again, it may well be that the author of *sūtra* Q2, having got an accurate enough three-term formula for reasonably small $r/(2m)$, was daunted by the task of expressing in verse the fourth term, a fairly complicated rational expression involving two variables (m and r) and several numbers, rather than by mathematical difficulty. Not every mathematical versifier is a Śaṅkara Vāriyar.

With reasonable confidence, we may conclude from all this that the idea of recursion as captured in the method of successive refining has been a constant of Indian mathematical thinking and technique over the two millennia separating *YB* from the late vedic era. The earliest *Śulbasūtra*s were probably assembled some time after the compilation of the particular version of the *Yajurveda* containing the famous list of powers of 10 and the knowledge they display may well be contemporaneous with, or even anterior to, that text. As for Bakhshali, that comes, very likely, soon after those number-happy centuries when naming and listing impressively large numbers had become an end in itself. (A case can me made for its date to be somewhat earlier than Hayashi's, on the basis of what is reasonably certain about the political history of northwest India). The intervening period is surely one in which Indian thinkers, in addition to mastering arithmetic, lost their fear of infinity as the philosophical and canonical (including the astronomical) literature of Jainism, Buddhism and the Hinduism of the time demonstrates. From the recursive infinitude of numbers-with-a-base, it is a natural step to power series, a step taken also by Newton in his turn and, more generally, to the idea that a recursive computation need not end for it to make sense, that a well-defined quantity such as a number or the length of an arc can be expressed meaningfully in an unending series of terms. And the liberation probably came early – it is notable that all three of the *Śulba* manuals that give the approximate value of $\sqrt{2}$ end their enunciation of the formula with the phrase *saviśeṣaḥ*, "with remainder".

Thus, born out of the systematics of the decimal number nomenclature and marking its first success as a clever scheme of approximation in arithmetical problems, recursive thinking acquired such depth and richness with the evolution of mathematical ideas that, by the time of Mādhavan, it would seem that mathematics in every one of its aspects, conceptual, computational and logical, was almost inconceivable except in fundamentally recursive terms. Among the milestones which mark this journey three are particularly significant: the decimal numbers and the development of arithmetic of course, Āryabhaṭa's trigonometry and the sine table, and the invention and elaboration of calculus. Indeed a reading of *Yuktibhāṣā* makes one wonder whether Mādhavan's calculus would at all have been possible without the power and subtlety that these methods had by then acquired.

This work has been made possible by a Senior Award of the Homi Bhabha Fellowships Council and the hospitality of the IUCAA, Pune, to both of which institutions I express my thanks. From among friends and colleagues whose specific suggestions and clarifications have been of help, I mention particularly Bhagyashree Bavare, Chandrashekhar Cowsik, Roddam Narasimha and Nisha Yadav. David Mumford's comments on earlier drafts have significantly improved this final version. But, above all, I ackowledge his and Frits Staal's generous encouragement by way of queries, exchanges of ideas, etc. over the years.

Bibliography

[1] K. V. Sarma (Ed.), *Tantrasaṅgraha of Nīlakaṇṭha Somayāji with Yuktiīpikā and Laghuvivṛti of Śaṅkara* (Hoshiarpur, Vishveshvaranand Vishvabandhu Institute of Sanskrit and Indological Studies, 1977).

[2] M. S. Sriram, K. Ramasubramanian and M. D. Srinivas (Eds.), *500 Years of Tantrasangraha* (Shimla, Indian Institute of Advanced Study, 2002).

[3] Roddam Narasimha, "The Yukti of Classical Indian Astronomical Science" in B.V.Sreekantan (Ed.), *Foundations of Science* (New Delhi, PHISPC, 2007).

[4] K. V. Sarma, *Gaṇita-Yukti-Bhāṣā of Jyeṭhadeva* Vol.1 (New Delhi, Hindustan Book Agency, 2008) (cited as [Sarma]).

[5] Rama Varma (Maru) Tampuran and A. R. Akhilesvara Ayyar, *Yuktibhāṣā, Part 1 (General Mathematics)* (in Malayalam) (Trichur, Mangalodayam Ltd. 1948) (cited as [TA]).

[6] T. A. Sarasvati Amma, *Geometry in Ancient and Medieval India* (Delhi, Motilal Banarsidass Publishers, 1979, 1999).

[7] Frits Staal, *Ritual and Mantras: Rules Without Meaning* (Indian edition: Delhi, Motilal Banarsidass Publishers, 1996).

[8] Frits Staal, *Agni: The Vedic Ritual of the Fire Altar* (Indian edition: Delhi, Motilal Banarsidass Publishers, 2001).

[9] Bibhutibhushan Datta and Avadhesh Narayan Singh, *History of Hindu Mathematics* (Reprint: Delhi, Bharatiya Kala Prakashan, 2001).

[10] Frits Staal, "On the Origins of Zero", this volume.

[11] David Mumford, "What's so baffling about negative numbers? – a cross-cultural comparison", this volume.

[12] Frits Staal, "Artificial Languages Between Innate Faculties", J. Ind. Philos. **35**, 577 (2007).

[13] Frits Staal, *Discovering the Vedas* (New Delhi, Penguin Books, 2008).

[14] Georges Ifrah, *Histoire universelle des chiffres* (Paris, Robert Laffont, 1994).

[15] Takao Hayashi, *The Bakhshali Manuscript* (Groningen, Egbert Forsten, 1995).

[16] P. P. Divakaran, "The First textbook of Calculus – *Yuktibhāṣā*", J. Ind. Philos. **35**, 417 (2007).

[17] K. Ramasubramanian, M. D. Srinivas and M. S. Sriram, Explanatory notes to chapters 1–7 of [4] (cited as [RSS]).

[18] D. T. Whiteside (Ed.), *The Mathematical Papers of Isaac Newton* (Cambridge, The University Press, 1968).

[19] H. W. Turnbull (Ed.), *The Correspondence of Isaac Newton* (Cambridge, The University Press, 1960).

[20] John Playfair, "Observations on the Trigonometrical Tables of the Brahmins", Trans. Royal Soc. Edinburgh **4**, 84 (1795).

[21] Agathe Keller, *Expounding the Mathematical Seed* (Basel, Birkhäuser Verlag, 2006).

[22] S. Sinha, N. Yadav and M. Vahia, work in preparation.

[23] Walter A. Fairservis Jr., "The Script of the Indus Valley Civilization", Scient. American **248**, 59 (1983).

[24] Amartya Sen, *The Argumentative Indian* (New Delhi, Penguin Books, 2006).

[25] Frits Staal, "The Sound Pattern of Sanskrit in Asia", lecture at the International conference on Sanskrit in Asia (Bangkok, 2005), available at http://fritsstaal.googlepages.com/

[26] C. T. Rajagopal and M. S. Rangachari, "On Medieval Kerala Mathematics", Arch. Hist. Exact Sciences **35**(2), 91 (1986).

[27] Ebenezer Burgess, *The Sûryasiddhânta* (Reprint: Delhi, Motilal Banarsidass Publishers, 1989).

[28] Kripa Shankar Shukla, *Āryabhaṭīya of Āryabhaṭa* (New Delhi, Indian National Science Academy, 1976).

[29] Kim Plofker, "The 'Error' in the Indian 'Taylor Series Approximation' to the Sine", Historia Mathematica **28**, 283 (2001).

[30] Kim Plofker, "Relations between Approximations to the Sine in Kerala Mathematics", in G.Emch, R.Sridharan and M.D.Srnivas (Eds.), *Contributions to the History of Indian Mathematics* (New Delhi, Hindustan Book Agency, 2005).

[31] S. N. Sen and A. K. Bag, *The Śulbasūtras of Baudhāyana, Āpastamba, Kātyāyana and Mānava* (New Delhi, Indian National Science Academy, 1983).

[32] M. N. Channabasappa, "On the Square Root Formula in the Bakhshali Manuscript", Ind. J. Hist. Science **11**, 112 (1976).

P. P. Divakaran

Inter-University Centre for Astronomy and Astrophysics, Pune

Planetary and Lunar Models in *Tantrasaṅgraha* (c.1500) and *Gaṇita-Yukti-bhāṣā* (c.1530)

M. S. Sriram

1. Introduction

Āryabhaṭīya (c.499) is the first extant text on mathematical astronomy in India to discuss the explicit algorithms for the computation of planetary positions [1]. Eccentric/epicyclic models are somewhat implicit in these algorithms in *Āryabhaṭīya*, and in the texts which followed. In contrast, the planetary theory in Ptolemy's *Almagest* [2] and the later Islamic works which were influenced by Ptolemy, as also the Copernicus's *de Revolutionibus* [3], are based on explicit geometrical models . In essence, the motion of any planet is viewed as a combination of uniform circular motions in these models. No such rigid principles are insisted upon in *Āryabhaṭīya* or even in the later Indian texts. Though most texts do not discuss the geometrical models at all, there are exceptions. Nīlakaṇṭha Somayājī who introduced a major revision of the traditional Indian planetary theory in his *Tantrasaṅgraha* [4,5,6], outlined a geometrical model of planetary motion in his other works, *Golasāra*, *Siddhāntadarpaṇa* and *Āryabhaṭīya-bhāṣya*. In this model, planets move in eccentric orbits inclined to the ecliptic around the mean Sun, which in turn goes aound the earth [7,8]. We have a detailed account of this model in *Gaṇita-Yukti-bhāṣā*, popularly known as *Yukti-bhāṣā* (c.1530), authored by Jyeṣṭhadeva [9]. Here onwards, [9] will be indicated by just *Gaṇita-Yukti-bhāṣā* in all the specific references to the passages in it.

Apart from discussing the geometrical picture of planetary motion, *Yukti-bhāṣā* also presents an interesting geometrical model of lunar motion while discussing the so called '*dvitīya - sphuṭa*' or the 'second correction'. This correction termed 'Evection' was first introduced by Ptolemy in his *Almagest*, where he proposed a geometrical model for Moon's longitude, which incorporates evection, apart from the correction due to the eccentricity of the orbit. Ptolemy's lunar model had a glaring defect as the ratio of the maximum and minimum apparent angular diameters of the Moon would be around 1.92 in this model, which is patently absurd. This deficiency

of Ptolemy's lunar theory had been rectified in some later Islamic works which otherwise follow the Ptolemaic framework, as for instance in Ibn al-Shāṭir's (1304-1375/6 CE) lunar theory,which seems to have been adopted by Copernicus also, except for small change in parameters [10]. In the Indian tradition, the second correction to the Moon makes its first appearence in *Laghumānasa* of *Mañjulācārya* (c.932) who essentially gave an analytical form for it[1][11]. *Tantrasaṅgraha* also gives the algorithm for the second correction which has the same form as in *Laghumānasa*, but goes ahead to give the formula for the true distance of the Moon, incorporating this. We have an explcit geometrical model for the same in *Yukti-bhāṣā*. This model does not show any problem of unresonable variation in the earth-Moon distance that is associated with Ptolemy's lunar theory. The corrections to Moon's longitude as given in the lunar theories of both *Tantrasaṅgraha*/*Yukti-bhāṣā* and Ibn al-Shāṭir compare reasonably with the modern expression for the two corrections, though the geometrical models are very different. Also, the patterns of variation of the lunar distance are different in the two models.

In section 2, we discuss the salient features of planetary computations in Indian *siddhānta*s begining with *Āryabhaṭīya*. In section 3, we provide a brief outline of the contents of *Tantrasaṅgraha* and *Yukti-bhāṣā*. The planetary model in *Yukti-bhāṣā* is described in detail in section 4. For the sake of comparison, the geocentric position of a planet in Kepler's model is discussed in Appedix 1. In section 5, we discuss the lunar theory of *Yukti-bhāṣā* and briefly compare it with the lunar thories of Ptolemy and Ibn al-Shāṭir which are outlined in Appendices 2 and 3 respectively.

2. Planetary Computations in Indian *Siddhāntas*

The *Siddhānta* period in Indian astronomy [12] begins with *Āryabhaṭīya* which was composed in AD 499. There were indeed earlier *siddhānta*s which have been quoted by the later astronomers but are not available today. Mathematics developed with astronomy in India. *Siddhānta*s contain correct mathematical solutions to astronomical problems, namely calculation of positions of planets (Sun, Moon, Mercury, Venus, Mars, Jupiter and Saturn), and diurnal problems involving spherical astronomy, and eclipses. The diurnal problems include, finding north-south directions, latitude of a place, Sun's diurnal path, its declination, determination of the times of sunrise/sunset, measurement of time (from shadow), relations among various celestial coordinates, calculation of *lagna* (point on the ecliptic which is on the horizon at any time), etc. The treatment of eclipses includes the calculation of the instant of conjunction, the two half-durations (from the first contact to the middle,

[1]Yallayya in his commentary on *Laghumānasa* mentions that this correction appears earlier in *Vaṭeśvarasiddhānta*(c.904) of Vaṭeśvara. However, the currently available version of the latter work does not include the second lunar correction.

Table 1: Planetary revolutions in a *Mahāyuga* as given in *Āryabhaṭīya*, and the inferred sidereal periods. * The revolution numbers of Mercury and Venus refer to their *śīghroccas* (heliocentric values).

Planet	No. of Revolutions in a *Mahāyuga*	Period of Revolution in days (Sidereal period)	Modern value
Sun	43,20,000	365.25868	365.25636
Moon	5,77,53,336	27.32167	27.32166
Moon's apogee	4,88,219	3231.98708	3232.37543
Moon's nodes	2,32,226	6794.74951	6793.39108
Mercury*	1,79,37,020	87.96988	87.96930
Venus*	70,22,388	224.69814	224.70080
Mars	22,96,824	686.99974	686.97970
Jupiter	3,64,224	4332.27217	4332.58870
Saturn	1,46,564	10766.06465	10759.20100

and from the middle to the last contact), duration of totality and the magnitude of the eclipse, and so on.

Āryabhaṭīya is the earliest available *siddhānta* text in India, which contains a systematic account of most of the traditional astronomical problems [1]. It is mentioned in the text itself that it was composed 3600 years after the beginning of *Kaliyuga* corresponding to AD 499 when it is stated that the author was 23 years in age. Āryabhaṭa composed this work in *Kusumapura*, which is the same as Paṭalīputra (near modern Patna). From *Āryabhaṭīya* onwards, one considers a *Mahāyuga* of 43,20,000 years. The number of revolutions in the stellar background made by the planets in the *Mahāyuga* are given in all the texts (see Table 1).

In the case of Mercury and Venus, the texts give the revolution numbers of their *śīghrocca*-s, which are nothing but their heliocentric revolution numbers. The number of civil days in a *Mahāyuga* known as the *Yugasāvanadina* is also specified. In *Āryabhaṭīya*, its value is given as 1,57,79,17,500.

From the above data, the mean longitudes of the planets can be calculated for any time. Normally, it is assumed that the mean longitudes are zero at the beginning of *kaliyuga*. In *Āryabhaṭīya*, this is taken to be the mean sunrise at Ujjain on February 18, BC 3102.

Now, the apparent motion of the Sun, Moon and planets in the background of stars is not uniform. The planets (including the earth) and the Moon move in elliptical orbits around the Sun, and the earth respectively. The ellipticity of the orbit gives rise to the non-uniform nature of the motion. This is essentially Kepler's picture. In the geocentric frame of reference, the Sun moves around the earth, and the geocentric longitude of the Sun would be the heliocentric longitude of the earth increased by 180°. The geocentric longitudes of the planets would be related to their heliocentric longitudes and the geocentric longitude of the Sun.

Even in the ancient times, the predictions of planetary positions were fairly accurate. This means that the ancient models would have in some sense been approximations of the Kepler model in the geocentric frame [13]. In the Indian planetary theories, two corrections were applied to the mean planet to obtain the 'true' geocentric longitudes, and they are described below:

(1) *Manda-saṃskāra*: This takes into account the non-uniformity of motion resulting from the eccentricity of the planet's orbit. An epicyclic or eccentric model is used to obtain the *manda*-correction. This corresponds to the 'Equation of centre' in modern terminology. The *manda*-corrected mean planet is called *mandasphuṭa-graha* or simply *mandasphuṭa*. This is the true geocentric longitude in the case of the Sun. For the Moon, a second correction has been specified in some later texts to obtain the geocentric longitude, which will be discussed later. In the case of the planets called *tarāgraha*-s (traditionally, only Mercury, Venus, Mars, Jupiter and Saturn), the *mandasphuṭa* is the true heliocentric longitude.

(2) *Śīghra-saṃskāra*: For the planets, one more correction, namely *śīghra*, has to be applied to obtain the true longitude called *sphuṭagraha*. The *śīghra-saṃskāra* converts the heliocentric longitudes of the planets to geocentric longitudes.

In *Āryabhaṭīya*, the above two corrections have been discussed clearly. The planetary model described by Āryabhaṭa roughly amounts to the planets orbiting around the Sun in eccentric orbits, with the Sun itself orbiting around the earth, though Āryabhaṭa does not state it as such. We discuss the Indian planetary models, in particular the traditional confusion in taking the mean Sun as the mean planet in the case of the interior planets, Mercury and Venus, later. Significantly, the picture of latitudes of planets is broadly correct in *Āryabhaṭīya*.

We may mention here some of the later astronomers and their major works [12]: Varāhamihira (b. AD 505) (*Pañcasiddhāntikā*), Bhāskara I (7th century) (*Mahābhāskarīya*, *Āryabhaṭīya-bhāṣya*), Brahmagupta (7th century) (*Brāhmasphuṭasiddhānta*), Lalla (8th-9th century) (Śiṣyadhīvṛddhida-tantra), Vaṭeśvara (9th century) (*Vaṭeśvarasiddhānta*), Mañjulācārya (10th century) (*Laghumānasa*), Śrīpati (11th century) (*Siddhāntaśekhara*), Bhāskarācārya II (b. AD 1114) (*Siddhāntaśiromaṇi* with *Vāsanābhāṣya*). It is in *Laghumānasa* that the second correction for the Moon corresponding to 'Evection' makes its first appearence in the Indian tradition. Bhāskara II has also included this correction for the Moon. This correction arises due to the influence of the Sun on the lunar orbit.

Siddhāntaśiromaṇi is a compendium which contains most of the results known to Indian astronomers at that time. After Bhāskara II, there was very significant work in astronomy in Kerala during 14th - 17th centuries, which also lead to

some major innovations in planetary theory. We will discuss them in the next section.

3. *Tantrasaṅgraha* and *Yukti-bhāṣā*

Kerala was a stronghold of Āryabhaṭan school of astronomy [14]. *Āryabhaṭīya* became popular in Kerala from early times. Mādhava of Sangamagrama (1340-1425) seems to have been the founder of the Kerala school. His known works are few in number. However, most of the distinctive contributions of the Kerala school in mathematical analysis, as well as in planetary computations are invariably attributed to him by the later astronomers from Kerala. Parameśvara of Vaṭasseri (c.1360-1455), Nīlakaṇṭha Somayājī or Somasutvan of Trikkantiyur (c.1444-1550) and Jyeṣṭhadeva of *Parakroḍa* or *Paraṅṅoṭṭu* family (c.1500-1610), are some of the major figures of this school. Emphasizing the need for revising the planetary parameters through observations, Parameśvara introduced the *Dṛggaṇita* system. Apart from *Dṛggaṇita*, his other major works are *Bhaṭadīpikā* which is a commentary on *Āryabhaṭīya*, *Siddhāntadīpikā*, and *Grahaṇamaṇḍana* on eclipses. He was one of the first astronomers to discuss the geometrical model of planetary motion in some detail.

Tantrasaṅgraha, composed in AD 1500, is Nīlakaṇṭha's major work. *Āryabhaṭīya-bhāṣya* composed by him late in his life is perhaps the most elaborate commentary on *Āryabhaṭīya*, and is yet to be translated and studied in depth. *Jyotirmīmāṃsā*, *Siddhānta-darpaṇa* and *Golasāra* are some of his other important works. Jyeṣṭhadeva is the author of *Gaṇita-yukti-bhāṣā*. He authored one more work called *Dṛkkarṇa*.

Tantrasaṅgraha

Tantrasaṅgraha is a comprehensive text which discusses all aspects of mathematical astronomy such as the computation of longitudes and latitudes of planets, various diurnal problems, determination of time, eclipses, visibility of planets etc. However, explanations are not provided, save on some rare occasions, as this belongs to the *Tantra* class of texts which are intended to be computational in nature. The explanations of the algorithms are to be found in *Yukti-bhāṣā*, and in the two commentaries of the text, namely *Laghuvivṛti* and *Yukti-dīpikā*, both by Śaṅkara Vāriyar of Trikkutaveli (c.1500-1562). In what follows, we summarise those portions of *Tantrasaṅgraha* which deal with planetary and lunar theory.

The revolution numbers of the Sun, Moon and the five planets given in this text differ little from the Āryabhaṭan values summarised in Table 1. The *yugasāvanadina* or the number of civil days in a *yuga* is also the same as in *Āryabhaṭīya*. However, it

is noteworthy that while specifying the number of revolutions of the interior planets, Mercury and Venus, the word *svaparyayāḥ* is used, which clearly means that the revolution numbers given refer to their own revolutions, and not to their *śīghroccas*, as specified in the earlier texts. This is of great significance in the computation of longitudes of the inner planets, and will be explained later.

The revolution numbers of the apsides and nodes of the planets which are needed for calculating the true longitudes and latitudes are also given. The mean longitudes at the begining of *kaliyuga* are not assumed to be zero, as in *Āryabhaṭīya*, and their corrected values are specified. The mean longitude of a planet can be calculated at any given time, given these epochal values, the revolution numbers and the *yugasāvanadina*.

In the earlier Indian texts, for the interior planets Mercury and Venus, the equation-of-centre-correction was wrongly applied to the mean Sun, instead of the mean heliocentric planet. This is true of Ptolemy's *Almagest* also, and clearly this correction does not correspond to anything physical. It was Nīlakaṇṭha who set this right in *Tantrasaṅgraha*. Here the *manda* correction is applied to the mean heliocentric planet for the interior planets also, just as in the case of exterior planets. This departure, as well as a clear analysis of the latitudinal motion, led Nīlakaṇṭha to propose a geometrical picture of planetary motion in his other works. According to this, the planets move in eccentric orbits around the mean Sun, which itself moves around the earth [7,8].

The actual distances of the Sun and the Moon from the centre of the earth play an important role in eclipse calculations, as these determine the apparent sizes of the solar and lunar discs, as well as their parallaxes. This requires accurate computations of the positions of the Sun and the Moon. It is in this context that the second correction for the Moon is discussed in *Tantrasaṅgraha*. There is a similar correction for the Sun ,which does not alter its longitude, but modifies its distance very marginally. The mean distances are specified in the text. The true distance called *dvitīya-sphuṭa-yojana-karṇa* is to be computed by including the *manda* correction and the above second correction.

Yukti-bhāṣā

It is very significant that *Yukti-bhāṣā* is written in Malayalam, the local language of Kerala. It is really a textbook for students as it systematically develops the relevant mathematics and astronomy, and is self-contained. The basics are set forth at the beginning of each topic and all the results which follow are proved. The main concepts are often repeated for emphasis. Alternate methods for important formulae are given, wherever possible. There are no equations as we write now, nor do we find any diagrams in the available manuscripts. However, from the detailed descriptions given in the text, one can write down the equations, as also draw the figures without ambiguity.

Yukti-bhāṣā can be naturally divided into mathematics and astronomy portions. The mathematics part is really an independent treatise. A substantial part of mathematics known to Indians at that time is included and explained in the work. The famous results on the infinite series for π, inverse tangent, sine and cosine functions are all proved here. It also includes a derivation of the surface area and volume of a sphere using methods which amount to carrying out integrations.

The astronomy part is very comprehensive, and includes almost all that was known in India then, save some topics like instruments. The author's stated aim is to explain all the algorithms in *Tantrasaṅgraha*.

The first chapter in the astronomy part is on planetary theory. The epicyclic and eccentric theories are described for both *manda-saṃskāra* (equation of centre) and *śīghra-saṃskāra* (conversion from heliocentric to geocentric coordinates). The correct formulation of the equation of centre for interior planets is presented and the way it differs from the earlier formulations is noted. The geometrical model of Nīlakaṇṭha is described in detail. The computation of geocentric latitudes is also described. The orbit of a planet has to be projected onto the plane of the ecliptic to obtain the longitude. This is explained. There is an interesting discussion on how one can obtain the geocentric coordinates of a celestial body, given the luni-centric coordinates. This indicates how comfortable Jyeṣṭhadeva was with coordinate transformations while going from one frame of reference to another, in the context of astronomy.

The second correction for the Moon is dealt with in a later chapter. The *Tantrasaṅgraha* results for this correction as well as the distance of the Moon are explained with a geometrical model.

4. Planetary Model in *Yukti-bhāṣā*

Manda-Saṃskāra- Correction Due to Eccentricity of the Orbit in Indian Planetary Models

We consider the explicit geometrical construction for computing the correction due to eccentricity described in *Yukti-bhāṣā* [15]. Essentially the same form of the correction term has been described right from *Āryabhaṭīya* down to *Tantrasaṅgraha*. The speciality of *Yukti-bhāṣā* lies in its detailed geometrical description.

In Figure 1, the dashed circle, with O as the centre is the deferent circle, termed '*kakṣyāvṛtta*'. Its radius is '*trijyā*', denoted by R. In the Indian texts, the distances in diagrams associated with planetary motion are expressed in minutes. Normally, the circumference of the deferent circle is taken to be 21,600 minutes, and thus the radius R would be the number of minutes in a radian, which is nearly $3438'$. Throughout this article, this is the value of R, wherever it appears. A represents the direction of the apside ('*mandocca*') and let $\Gamma\hat{O}A = \varpi$. The mean planet P_0 moves along the deferent circle uniformly. The mean longitude is called '*madhyamagraha*' and is

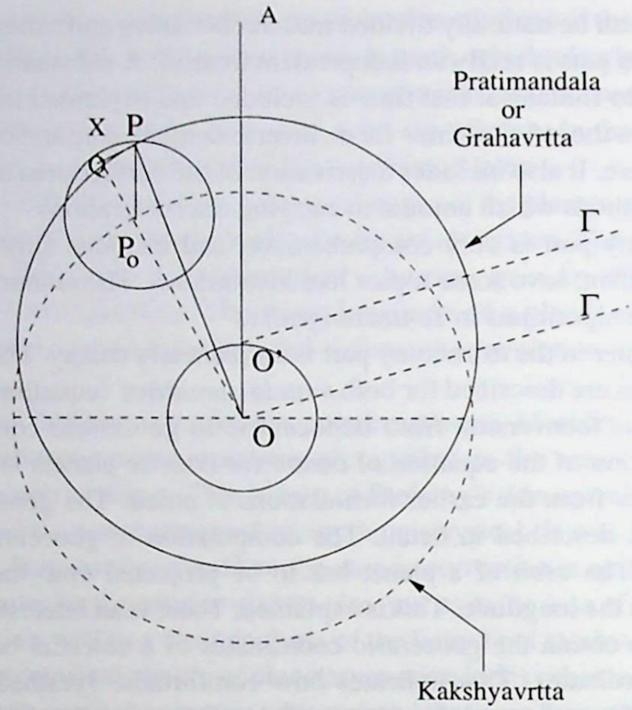


Figure 1: *Manda-saṃskāra* or the correction due to eccentricity of the orbit.

given by $\theta_M = \Gamma\hat{O}P_0$.[2] Construct a circle of radius r with P_0 as the centre. This is the '*manda*'-epicycle ('*manda-nīcocca-vṛtta*' or '*manda-vṛtta*') on which the '*manda*'-corrected planet ('*mandasphuṭa*') is located. Draw a line from P_0 parallel to OA, which intersects the epicycle at P. This is the location of the *manda*-corrected planet whose longitude is given by $\theta_{MS} = \Gamma\hat{O}P$. This is the epicycle model.

Alternatively, locate O' on OA such that $OO' = r$. The circle with O as the centre and r as the radius is also known as *manda*-epicycle. Draw a circle with radius R with O' as the centre. This is the eccentric circle, called the '*pratimaṇḍala*'. It is easy to see that the *manda*-corrected planet , P is located on this eccentric circle such that $\Gamma\hat{O}'P = \Gamma\hat{O}P_0 = \theta_M$. That is, the planet moves uniformly around O', and not around O with respect to which the longitude is measured. This is the 'eccentric circle' model. Clearly, it is equivalent to the 'epicycle model'.

---

[2]In this article, all the longitudes and latitudes are with respect to the ecliptic. For calculating the planetary positions, the zero point for the longitudes is the '*mesādi*' or the 'first point of Aries', denoted by Γ in this article. In the Indian texts, the zodiacal signs are fixed in the background of stars and the corresponding longitudes are the *nirayaṇa* ones. The 'tropical longitudes' which are qualified as *sāyana*, corresponding to the vernal equinox as the zero point, are also used particularly in the treatment of diurnal problems. The difference between these two longitudes is due to the 'precession of equinoxes', and does not affect any of the arguments in the present article.

Draw PQ perpendicular to (extended) OP_0. Now, $PP_0 = r$. $P\hat{P}_0Q = \Gamma\hat{O}P_0 - \Gamma\hat{O}A = \theta_M - \varpi = M$ is the anomaly or *manda-kendra*. and $P\hat{O}P_0 = \Gamma\hat{O}P_0 - \Gamma\hat{O}P = \theta_M - \theta_{MS}$. Now

$$PQ = r\sin(P\hat{P}_0Q) = OP\sin(P\hat{O}P_0). \tag{1}$$

$K = OP$ is the '*manda*'-hypotenuse, called '*mandakarṇa*'. It is easily seen that

$$K = [(R + r\cos M)^2 + r^2\sin^2 M]^{1/2}. \tag{2}$$

Yukti-bhāṣā follows the standard rule adopted in most texts of Indian astronomy (such as *Āryabhaṭīya*, *Brāhmasphuṭasiddhānta*, *Tantrasaṅgraha*, etc.), that r is not a constant but proportional K, such that

$$\frac{r}{K} = \frac{r_0}{R}, \tag{3}$$

where r_0 is prescribed in the text for each planet. r_0 is essentially the mean radius of the *manda*-epicycle, and could itself depend on the anomaly for some planets. Now K depends on r, and r is proportional to K. In *Tantrasaṅgraha* and *Yukti-bhāṣā*, an iterative procedure is given to find K. K can be found non-iteratively also, using an elegant geometrical method [16]. The exact expression for the *manda*-hypotenuse, K is:

$$K = \frac{R^2}{\sqrt{R^2 - r_0^2\sin^2 M} - r_0\cos M}. \tag{4}$$

Thus K is determined in terms of known quantities. This exact expression is attributed to Mādhava by Nīlakaṇṭha. As r is not a constant, the planet's orbit is not a circle, but some kind of an oval.

It would be useful to summarise the definitions of various relevant quantities at this stage:

$$\begin{aligned}
\text{mean longitude}(\textit{madhyama-graha}) = \Gamma\hat{O}P_0 = \Gamma\hat{O}'P &= \theta_M,\\
\text{apside } (\textit{mandocca}) = \Gamma\hat{O}A &= \varpi,\\
\text{anomaly } (\textit{manda-kendra}) = A\hat{O}P_0 = P\hat{P}_0Q &= \theta_M - \varpi,\\
\textit{manda}\text{-hypotenuse } (\textit{manda-karṇa}), K &= OP,\\
\textit{manda}\text{-corrected longitude } (\textit{manda-sphuṭa}), \theta_{MS} &= \Gamma\hat{O}P,\\
\textit{manda}\text{-correction, } \textit{manda-phala} &= \theta_{MS} - \theta_M.
\end{aligned}$$

Taking (3) into account, (1) gives

$$\sin(\theta_{MS} - \theta_M) = -\frac{r_0 \sin(\theta_M - \varpi)}{R} = -\frac{r_0 \sin M}{R}. \quad (5)$$

Thus, one does not need to know the *manda*-hypotenuse in order to find the correction $\theta_{MS} - \theta_M$. As the mean longitude, θ_M which varies linearly with time can be readily found at any instant, and apside, ϖ is known, the '*manda*'-correction ('*mandaphala*'), $\theta_{MS} - \theta_M$ can be found from the above equation. Adding this to the mean longitude θ_M, we obtain the '*manda*'-corrected planet, θ_{MS}. This is the true longitude or the '*sphuṭagraha*' in the case of the Sun, and essentially so also for the Moon.

The *manda*-corrected planet is the true heliocentric longitude for the planets Mercury, Venus, Mars, Jupiter and Saturn called '*tārāgraha*'s, as we will see shortly. One more correction, namely '*śīghra*'-correction ('*śīghra-saṃskāra* ') has to be applied to them to obtain their geocentric longitudes.

When $\frac{r_0}{R} \ll 1$, $\sin(\theta_{MS} - \theta_M) \approx \theta_{MS} - \theta_M$ and (5) reduces to

$$\theta_{MS} - \theta_M = -\frac{r_0 \sin(\theta_M - \varpi)}{R} = -\frac{r_0 \sin M}{R}. \quad (6)$$

Comparing this with (35) in Appendix 1, it can be seen that this has the same form as the equation of centre correct to lowest order in eccentricity in Kepler's model, with $\frac{r_0}{2R}$ playing the role of eccentricity, e. This equivalence is only to the first order in e. The equation of centre has different forms in higher orders of e in Kepler's model and Indian models.

In Ptolemy's model for the equation of centre, the planet is taken to be moving on the 'eccentric circle' with O' as the centre alright, but it would be moving uniformly not with respect to O', but with respect to a point which is at a distance 2 $O\,O'$ from O in the direction of O' [17]. This point is called the 'equant'. Ptolemy's model also would be equivalent to Kepler's model to the first order in eccentricity.

Śīghra-saṃskāra: Conversion to the Geocentric Frame

Exterior Planets

The following explicit geometrical construction is described in *Yukti-bhāṣā* to obtain the geocentric longitude of an exterior planet [18]. The final formula is essentially the same as in the earlier texts.

In Figure 2, O is the centre of the earth ('*Bhagola-madhya*'). The '*śīghra*' epicycle ('*śīghra-nīcocca-vṛtta*', or *śīghra-vṛtta*') - is a circle with O as the centre, and whose radius, r_s is given. The '*śīghrocca*', S is located on this circle. It is also stated that *śīghrocca* is the mean Sun ('*āditya-madhyama*'). The *manda*-epicycle

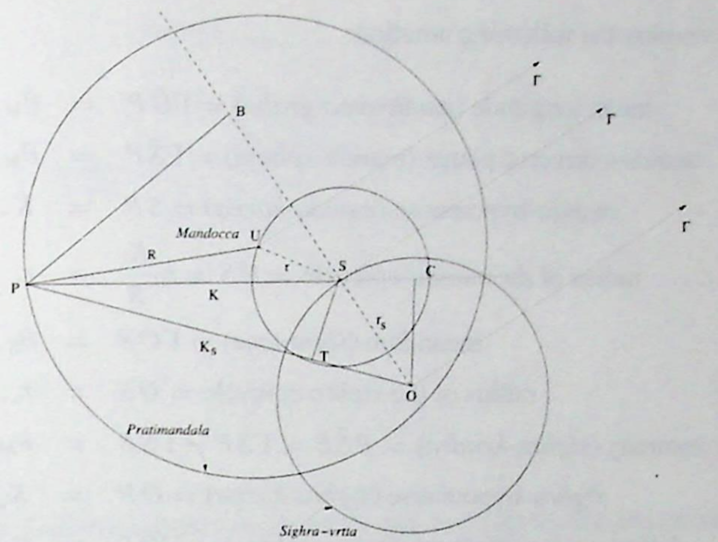


Figure 2: *Śīghra-saṃskāra* for exterior planets to obtain the geocentric longitude.

is a circle with the *śīghrocca* as the centre. The planet's apside (*mandocca*), U is located on this circle. The planet P is located on the eccentric circle which is centered at U. $K = SP$ is the *manda*-hypotenuse. $P\hat{S}\Gamma$ is the *manda*-corrected planet, which is the true heliocentric longitude. $P\hat{S}B$ which is the difference between this and *śīghrocca* is known as *śighra*-anomaly (*śīghra-kendra*). $\Gamma\hat{O}P$ is the '*śighra*'-corrected planet ('*śīghra-sphuṭa*'). The *śīghrocca*, S is the mean Sun, and the *śīghra*-epicycle represents its orbit around the earth, O. The *manda*-corrected planet is the true heliocentric longitude of the planet around mean Sun, and the *śīghra*-corrected planet, $\Gamma\hat{O}P$ is the geocentric longitude of the planet with respect to the earth, O. It is in this sense that the *śīghra-saṃskāra* is a coordinate transformation, which converts the heliocentric longitude to the geocentric longitude.

The *śīghra*-corrected planet is found in the same manner from the *manda*-corrected planet , as the latter is found from the mean planet. Thus it may be noted that, in the computation of the *śīghra*-corrected planet, the *śīghrocca* and the *manda*-hypotenuse will play the same roles as the apside and the the radius of the eccentric circle (R) did in the computation of the *manda*-corrected planet. The *śīghra*-hypotenuse, $K_s = OP$ can be determined in terms of $SP = K$. Apart from the similarities, there is one difference. In the *manda*-correction, we had noted that the radius of the epicycle r increases or decreases in the same way as K. In the *śīghra*-correction, the radius $OS = r_s$, does not vary with the *śīghra*-hypotenuse, K_s. To start with, both the mean radius r_0 of the *manda*-epicycle and the radius r_s of the *śīghra*-epicycle are specified in the measure of the radius of the eccentric circle, being *trijyā* or $R = 3438'$.

We now employ the following notation:

$$\begin{aligned}
\text{mean longitude } (\textit{madhyama-graha}) = \Gamma\hat{U}P &= \theta_M, \\
\textit{manda}\text{-corrected planet } (\textit{manda-sphuṭa}) = \Gamma\hat{S}P &= \theta_{MS}, \\
\textit{manda}\text{-hypotenuse } (\textit{manda-karṇa}) = SP &= K, \\
\text{radius of the } \textit{manda}\text{-epicycle} = US = r_0\frac{K}{R} &= r, \\
\text{mean Sun } (\textit{śīghrocca}) = \Gamma\hat{O}S &= \theta_S, \\
\text{radius of the } \textit{śīghra} \text{ epicycle} = OS &= r_s, \\
\textit{śīghra}\text{-anomaly } (\textit{síghra-kendra}) = P\hat{S}B = \Gamma\hat{S}P - \Gamma\hat{S}B &= \theta_{MS} - \theta_S, \\
\textit{śīghra}\text{-hypotenuse } (\textit{śīghra-karṇa}) = OP &= K_s, \qquad (7) \\
\text{śīghra-corrected planet } (\textit{śīghra-sphuṭa}) = \Gamma\hat{O}P &= \theta_g. \qquad (8)
\end{aligned}$$

Here the *manda*-corrected planet, θ_{MS} and the *manda*-hypotenuse, K are determined in terms of the mean longitude (θ_M), the apside (ϖ), and the prescribed value of r_0, as described in the previous sub-section[3]. The following formula for θ_g can be easily derived from the geometrical construction in Figure 2:

$$\begin{aligned}
\sin(\theta_g - \theta_{MS}) &= -\frac{r_s \sin(\theta_{MS} - \theta_S)}{K_s}, \\
&= -\frac{r_s \sin(\theta_{MS} - \theta_S)}{[(K + r_s\cos(\theta_{MS} - \theta_S))^2 + r_s^2 \sin^2(\theta_{MS} - \theta_S)]^{1/2}}. \qquad (9)
\end{aligned}$$

This is the formula in earlier texts also, such as *Sūryasiddhānta*, *Mahābhāskarīya*, etc. This formula of *Yukti-bhāṣā* is exactly the same as would follow from the Keplerian model (see (41), of appendix 1) for an exterior planet, when θ_{MS} in the former is identified with the true heliocentric longitude in the latter (θ_h), and $\frac{r_s}{K}$ in the former is identified with $\frac{r}{R}$ in the latter. The identification of θ_{MS} with θ_h is clear enough, as θ_{MS} is the heliocentric mean planet to which the correction due to eccentricity has been added. There is a small error as θ_s is the mean Sun and not the true Sun. This error is present in Ptolemy's Almagest [2] also, and persists even in *de Revolutionibus* of Copernicus [3]. In Table 2, we compare the average values of $\frac{r_s}{R}$ which are the ratios of the earth-planet and sun-planet distances in the geometrical model with the ratios of the actual physical values of these. Clearly there seems to be reasonable agreement between these values.

---

[3] Note that the radius of the *manda*-epicycle, US itself is $r = r_0\frac{K}{R}$.

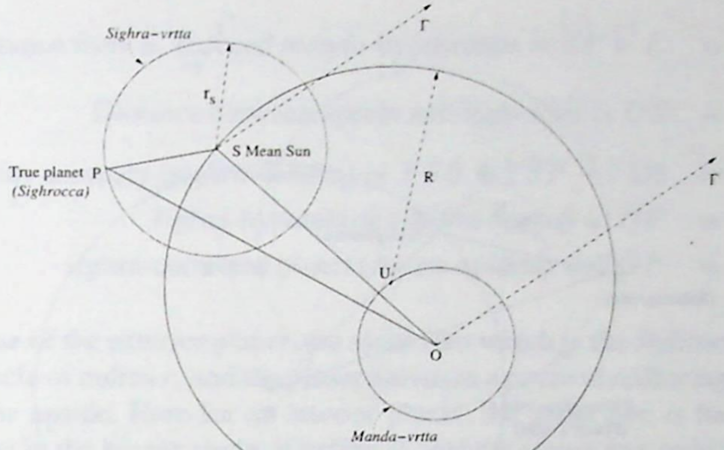


Figure 3: *Śīghra-saṃskāra* for interior planets to obtain the geocentric longitude, in *Āryabhaṭīya*.

Interior Planets

In both the Indian and Greek traditions, the mean Sun was taken to be the mean planet for the interior planets. This is due to the fact that the directions of Mercury and Venus are close to that of the Sun, as viewed from the earth. So, their geocentric longitudes would not be very different from the longitude of the Sun. The procedure for the computation of the geocentric longitude of an interior planet in *Āryabhaṭīya* and other texts prior to *Tantrasaṅgraha*, is represented geometrically in Figure 3. Here S is the mean Sun moving in a circle around U, which is the apside of the planet, where OU is equal to the radius of the *manda*-epicycle of the planet. Thus, S is nothing but the mean Sun, to which the *manda*-correction pertaining to the planet has been applied. P is the *śīghrocca* which is actually the mean heliocentric planet, moving on the *śīghra*-epicycle, which is the planet's orbit around S.[4] Then what is computed as the geocentric longitude of the inner planet in the earlier texts, is actually $\Gamma\hat{O}P$. This is obviously wrong. The *manda*-corrected planet $\Gamma\hat{O}S$ here does not correspond to anything physical. The same mistake was committed by Ptolemy in his *Almagest* [2] and even by Copernicus [3], who seems to have followed Ptolemy closely.

Nīlakaṇṭha Somayājī seems to have been the first astronomer who revised this picture in *Tantrasaṅgraha* and his other works [7,8]. His revised formulation is what is described in *Yukti-bhāṣā* [19] and depicted here in Figure 4. Nīlakaṇṭha has a

---

[4]In *Āryabhaṭīya* and other ancient texts, for the interior planets, the *śīghrocca* and the *śīghra*-epicycle represent entirely different things, when contrasted with the same objects for exterior planets. For the latter, the *śīghrocca* was the mean Sun and the *śīghra*-epicycle was mean Sun's orbit around the earth.

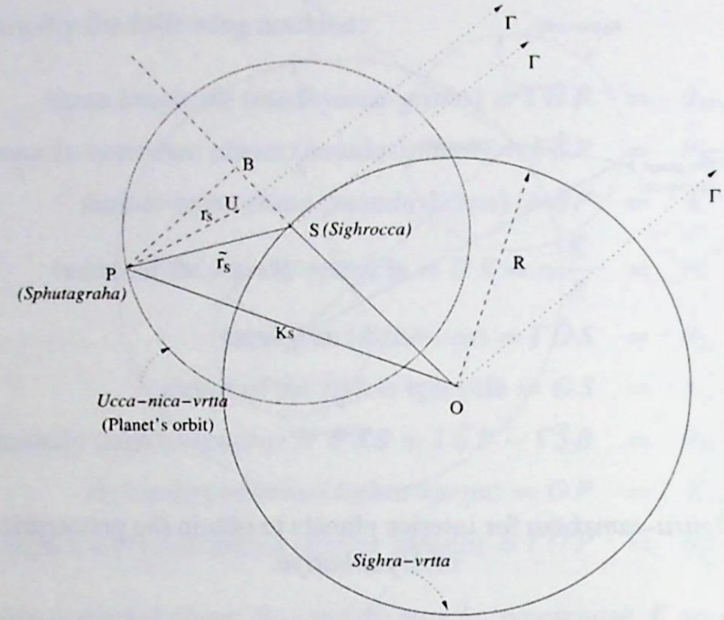


Figure 4: *Śīghra-saṃskāra* for interior planets in *Tantrasaṅgraha* and *Yukti-bhāṣā*.

unified formulation for the interior and exterior planets, where the true heliocentric longitude is converted to the geocentric longitude correctly. Here the mean Sun S is the *śīghrocca*, in the same way as for exterior planets. S moves on a circle of radius R with the centre of the earth, O as the centre. This is the *śīghra*-circle (*śīghra-vṛtta* or the *śīghra-nīcocca-vṛtta*) for the planet. The planet P moves on an eccentric orbit around S. The planet's distance from S, SP is the *manda*-hypotenuse with its value reduced from K to $\tilde{r}_s = K\frac{r_s}{R}$.[5] $\Gamma\hat{S}P$ is the *manda*-corrected planet, which is the longitude of P measured with respect to S. This is the true heliocentric longitude of the planet. In Figure 4, we have also shown the apside (*mandocca*), U. $UP = r_s$ and $US = \frac{r_s}{R}r$, which is the reduced radius of the *manda*-epicycle. The smaller circle in the figure is the instantaneous orbit of P around S, whose radius is $\tilde{r}_s$. $\Gamma\hat{O}P$ is the *sīghra*-corrected planet, which is the geocentric longitude . We now summarise the various relevant definitions:

$$\begin{aligned} \text{mean planet } (\textit{madhyama-graha}) = \Gamma\hat{U}P &= \theta_M, \\ \text{mean radius of the planet's orbit} = UP &= r_s, \\ \text{mean Sun } (\textit{śīghrocca}) = \Gamma\hat{O}S &= \theta_S, \\ \textit{manda}\text{-corrected planet } (\textit{manda-sphuṭa}) = \Gamma\hat{S}P &= \theta_{MS}, \end{aligned}$$

[5] The determination of the *manda*-hypotenuse, K has been explained above in the sub-section on *manda-saṃskāra* .

$$\text{Planet's distance from S, reduced } \textit{manda}\text{-hypotenuse} = SP = \tilde{r}_s = r_s \frac{K}{R},$$

$$\text{Distance between apside and } \textit{śīghrocca} = US = \frac{r_s}{R} r,$$

$$\text{śīghra-anomaly } (\textit{śīghra-kendra}) = P\hat{S}B = \Gamma\hat{S}P - \Gamma\hat{S}B = \theta_{MS} - \theta_S,$$

$$\text{śīghra-hypotenuse } (\textit{śīghra-karṇa}) = OP = K_s, \quad (10)$$

$$\text{śīghra-corrected planet } (\textit{śīghra-sphuṭa}) = \Gamma\hat{O}P = \theta_g. \quad (11)$$

In the case of the exterior planet, the mean Sun which is the *śīghrocca* moves in a smaller circle of radius r_s and the planet moves in a circle of radius equal to *trijyā*, R around the apside. Here for an interior planet, the mean Sun is the *śīghrocca*, which moves in the bigger circle of radius R, and the planet moves in an eccentric orbit, such that it is at a distance r_s from U, and $\tilde{r}_s$ from S. The *manda*-corrected planet , θ_{MS} is calculated from θ_M in the usual fashion. The *śīghra*-hypotenuse, K_s is given by

$$K_s = [(R + \tilde{r}_s \cos(\theta_{MS} - \theta_S))^2 + (\tilde{r}_s)^2 \sin^2(\theta_{MS} - \theta_S)]^{1/2}. \quad (12)$$

Then the geometry of Figure 4 implies the following formula for the true geocentric longitude, θ_g:

$$\begin{aligned} \sin(\theta_g - \theta_s) &= \frac{\tilde{r}_s \sin(\theta_{MS} - \theta_S)}{K_s}, \\ &= \frac{\tilde{r}_s \sin(\theta_{MS} - \theta_S)}{[(R + \tilde{r}_s \cos(\theta_{MS} - \theta_S))^2 + \tilde{r}_s^2 \sin^2(\theta_{MS} - \theta_s)]^{1/2}}. \quad (13) \end{aligned}$$

This is the same as (46) in appendix 1 for the geocentric longitude of an interior planet in Kepler's model, if we identify θ_{MS} here, with the true heliocentric longitude θ_h there, and $\frac{\tilde{r}_s}{R}$ in the above geometrical picture with $\frac{r}{R}$ in Kepler's model. Again, it is clear from the figure that θ_{MS} is the true heliocentric longitude, if we ignore the difference between the mean Sun and the true Sun. Ignoring the eccentricity of the inner planet's orbit, we compare the average values of $\frac{r_s}{R}$ which are the ratios of the Sun-planet and earth-planet distances in the *Tantrasaṅgraha* model, with the ratios of the modern physical values of these in Table 2. There is reasonable agreement, just as for the exterior planets.

Latitudes of Planets

Just as in the case of longitudes, we also have an unified theory of latitudes in *Yukti-bhāṣā* [20]. The geometrical picture for the latitudinal deflection is descibed as follows in the text [21].

Table 2: Comparison of $\frac{r_s}{R}$ in *Tantrasaṅgraha* for *śīghra-saṃskāra* with the modern values of the mean ratio of Earth-Sun and planet-Sun distances for exterior planets and the inverse ratio in the case of the interior planets.

Planet	*Tantrasaṅgraha* value	Modern value
Mercury	0.375	0.387
Venus	0.725	0.723
Mars	0.656	0.656
Jupiter	0.194	0.192
Saturn	0.106	0.105

"In the above set up, the *śīghra-vṛtta* is presumed with its centre at the centre of the *apakrama* circle [6]and its circumference along the *mārga* (in the plane) of the *apakrama* circle. It may be recalled that the *apakrama* circle near the centre is called *śīghrocca-nīca-vṛtta*. The size of the *śīghra-vṛtta*-s will be different for the different planets. That is all the difference (between the *śīghra-vṛtta*-s) and there is no difference in their placement as they are located the same way (i.e., with their centre at the centre of the *apakrama* circle and also lying in the same plane).

"Now, the *manda-nīcocca-vṛtta* is a circle having its centre on the circumference of the *śīghra-vṛtta* at the point where the mean Sun is. This is the case for all (the planets). The ascending node (*pāta*) has its motion along the circumference of the *manda-nīcocca-vṛtta* in the retrograde manner. The point in the *manda-nīcocca-vṛtta* where the *pāta* is, will touch the *apakrama-maṇḍala*. One half of the *manda-nīcocca-vṛtta*, commencing from the *pāta* will lie on the northern side of the *apakrama-maṇḍala*. Again, the point which is six signs away from the *pāta* will touch the *apakrama-maṇḍala*. The other half (of the *manda-nīcocca-vṛtta*) will lie on the southern side of the *apakrama-maṇḍala*. Here, that point, which is displaced maximum from the (plane of the) *apakrama-maṇḍala*, will indicate the maximum *vikṣepa* (*parama-vikṣepa*) of the planets in terms of the minutes of arc of their respective *mandocca-vṛtta*-s. Further, the plane of this *nīcocca-vṛtta* itself will be the plane of the *pratimaṇḍala*. Hence, the *pratimaṇḍala* too will be inclined towards the north and south from the plane of *apakrama-maṇḍala* in accordance with the *nīcocca-vṛtta*. The *manda-karṇa-vṛtta*[7] will also be inclined accordingly. Now, the *vikṣepa* has to be obtained from the *manda-sphuṭa*.

"Here, since the centre of the *manda-karṇa-vṛtta* is the same as the centre of the *mandocca-vṛtta* and since it will be inclined to the plane of *apakrama-maṇḍala*, south and north, accordingly as the *mandocca-vṛtta*, the maximum divergence of the circumference of the *manda-karṇa-vṛtta* from the plane of the *apakrama-maṇḍala* will be the maximum *vikṣepa* in the measure of the *manda-karṇa-vṛtta*. Hence, if

[6]ecliptic.

[7]A circle whose radius is the *manda*-hypotenuse.

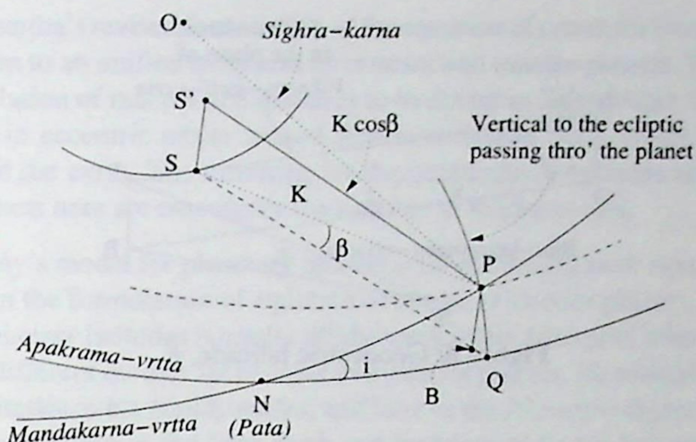


Figure 5: The latitudinal deflection (*vikṣepa*) of a planet .

the Rsine of the *manda-sphuṭa* minus *pāta* is multiplied by the maximum *vikṣepa* and divided by *trijyā*, the result will be the *iṣṭa-vikṣepa* of the planet on the *manda-karṇa-vṛtta*. This inclination (deflection from the ecliptic) is called *vikṣepa*."

This has the following implication for the geometrical model for the planets. In Figure 2 (for the exterior planets) and Figure 4 (for the interior planets), the mean Sun, S (*śighrocca*) moves on the *śīghra*-circle in the plane of the ecliptic (*apakrama-maṇḍala*), with the earth, O as the centre. The planet, P moves around S in an eccentric circle with the apside, U as the centre, in a plane which is inclined to the ecliptic. SU and UP are in this plane. First, let us consider the heliocentric longitude. In Figure 5, O is the earth and S is the Sun moving around the earth in the plane of the ecliptic. P is the planet moving in a plane inclined to the ecliptic, and $K = SP$ which lies in this plane is the *manda*-hypotenuse. The ecliptic, and the orbit of the planet intersect at N, which is the ascending node.

The latitude of the planet is $\beta = PQ$ which is the arc perpendicular to the ecliptic passing through P. PB is perpendicular to the plane of the ecliptic. Then the the distance of the planet from the plane of the ecliptic,termed '*vikṣepa*' is given to be

$$vik\d{s}epa = v = PB = K\sin\beta. \quad (14)$$

β itself is found from the following expression:

$$Rsin\beta = R\ \sin i\ \sin(\theta_{MS} - \theta_N),$$

where θ_{MS} is the true heliocentric longitude of the planet (*manda*-corrected planet), θ_N is the longitude of the ascending node of the planet, and i is the inclination of the orbit. Figure 6. shows the relation between the geocentric latitude $\tilde{\beta}$ and the heliocentric latitude β as explained in *Yukti-bhāṣā*. We have

$$PB = OP\sin\tilde{\beta} = SP\sin\beta, \quad (15)$$

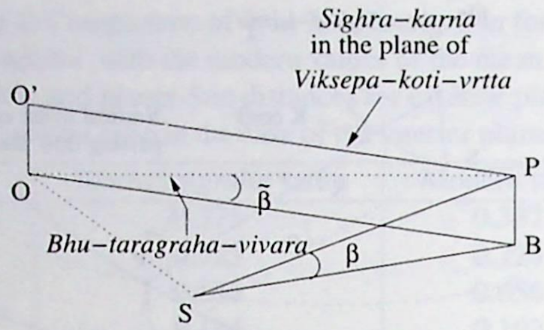


Figure 6: Geocentric latitude, $\tilde{\beta}$.

or,

$$\sin \tilde{\beta} = \frac{SP}{OP} \sin \beta. \tag{16}$$

When β and $\tilde{\beta}$ are small, the above relation reduces to

$$\tilde{\beta} = \beta . \frac{SP}{OP}. \tag{17}$$

OP is the earth-planet distance (*bhū-tārā-graha-vivara*), while SP is the Sun-planet distance. Equation (17) is exactly the same as the relation (51) in Appendix 1 on Kepler's model, relating the geocentric latitude β_E and the heliocentric latitude β_s. Here it may be pertinent to recall that the theory of latitude in Ptolemy's *Almagest* was totally unsatisfactory, as it took the planes of the orbits to be passing through the earth rather than the Sun [2]. The same was the problem with *de Revolutionibus* of Copernicus also [3].

The latitudinal deflection affects the longitude also. The *manda*-hypotenuse, $SP = K$, should be projected onto a plane passing through the planet, P which is parallel to the ecliptic. $S'P = K \cos \beta$, which is the projection is termed *vikṣepa-koṭi* (See Figs.(5) and (6)). It is pointed out that the geocentric longitude, *śīghra*-corrected planet, should be calculated taking the *vikṣepa-koṭi* as the *manda*-hypotenuse. The result is the true planet (*graha-sphuṭa*). This is the "reduction to the ecliptic" which was taken note of by some of the Islamic astronomers and later discussed by Tycho Brahe in late sixteenth century.

In summary, we may note the following important points as regards the development of planetary theory in Kerala texts:

1. The algorithms for the geocentric longitudes of planets given in *Āryabhaṭīya* and other ancient texts more or less simulated the results of the Kepler model, but for the fact that the equation of centre for interior planets was wrong.

2. Nīlakaṇṭha's revised formulation of the equation of centre for interior planets led him to an unified treatment of exterior and interior planets. The clearest articulation of this revised model is to be found in *Yukti-bhāṣā*. The planets move in eccentric orbits around *śīghrocca* (mean Sun), which itself goes around the earth. The formulae for the geocentric longitudes and latitudes of planets here are essentially the same as in Kepler model.

3. Ptolemy's model for planetary motion is more complicated. Apart from the error in the formulation of equation of centre of interior planets, the theory for planetary latitudes is totally off the mark in his *Almagest*, where Ptolemy gives different models for exterior and interior planets. He revised his model for latitudes in his *Handy tables*, and later in the *Planetary hypothesis* [22]. His formulation in the later work can be taken to be his last word on the subject. Here, in the case of the exterior planets, the eccentric (around the earth) is inclined to the ecliptic. The epicycle which represents the motion of the Sun is parallel to the ecliptic. The model for the interior planets is more complicated, but is a vast improvement over the same in *Almagest*. The final results for the latitudes in *Planetary hypothesis* are close to those following from the modern heliocentric theory, though they are complicated, and are not equivalent to Kepler's model in any simple manner.[8] In fact, the same is true with respect to the Copernican model also, revolutionary as it was in treating the earth also as a planet.

4. It should be mentioned that the geometrical model described in *Yukti-bhāṣā* is for each individual planet separately. It does not give an unified picture involving all the planets, to a single scale.

5. Lunar Theory in *Yukti-bhāṣā*

Yukti-bhāṣā describes a geometrical picture of the lunar motion which also incorporates the 'second correction' or the 'evection' correction [23].[9] This is indeed a very interesting geometrical picture, according to which the centre of the earth is not the same as that of the celestial sphere (*bhagola*), and the latter is in motion with respect to the earth in such a way that it is always along the earth-Sun direction or opposite to it, but at a variable distance. Now, the *manda*-corrected Moon, calculated using the procedure described earlier, is taken to be the true

[8] The *Handy tables* and *Planetary hypothesis* seem to have been totally ignored (at least as regards the theory of latitudes) by all the astronomers in the Greco-European and Islamic traditions till the time of Kepler [22].

[9] It may be noted that the geometrical picture of lunar motion presented in *Yukti-bhāṣā* is very different from the model suggested recently by Duke [24], as a possible geometrical picture to understand the second lunar correction discussed by Mañjulācārya in *Laghumānasa*.

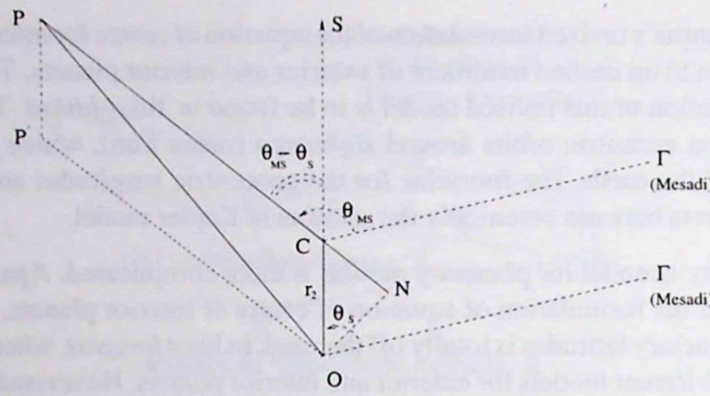


Figure 7: Geometrical model for the second correction to the Moon in *Yukti-bhāṣā*.

longitude with respect to the centre of the celestial sphere. A second correction has to be applied to the *manda*-corrected Moon to obtain the true longitude of the Moon with respect to the centre of the Earth, termed the 'second-corrected planet' ('*dvitīya-sphuṭa*').

The procedure for the second correction is similar to the calculation of the *manda*-correction with the centre of the celestial sphere serving as the equivalent of the apside, and called just '*ucca*', which is taken to be in the direction of the Sun . The distance between this and the centre of the Earth, which is the radius of the epicycle, is posited to be a continuously varying quantity given by $\frac{R}{2}\cos(\theta_S - \varpi)$ in *yojanā*[10]s, where θ_S and ϖ are the longitude of the Sun and the apogee of Moon (*candrocca*), respectively.[11] According to the Āryabhaṭan school, the mean distance between the Moon and the centre of the earth is $10R = 34380$ *yojanā*s. The actual distance between the Moon and the earth will be $10K$, where K is the *manda*-hypotenuse in minutes.

In what follows, we ignore Moon's latitude. In Figure 7, O is the centre of the Earth, separated from the centre of the earth (C) by a distance

$$r_2 = \frac{R}{2}\cos(\theta_S - \varpi) \qquad \text{(in } yojan\bar{a}\text{ s).} \tag{18}$$

Γ is the first point of Aries, and $\Gamma\hat{O}C = \theta_S$ (Sun's longitude). The Moon is at P. Hence $\Gamma\hat{C}P = \theta_{MS}$ (*manda*-corrected Moon). Here θ_{MS} incorporates the correction due to equation of centre (*manda-saṃskāra*), as described in section 4. The mean

[10] a unit of distance in Indian texts.

[11] When $\cos(\theta_S - \varpi)$ is negative, the centre of the *bhagola* is at a distance $|\frac{R}{2}\cos(\theta_S - \varpi)|$ from the centre of the earth, opposite to the direction of the Sun.

longitude, θ_M and the apogee ϖ are not shown in the figure. $CP = D_1 = 10K$, where K is the *manda*-hypotenuse in minutes. It is clear that $O\hat{C}N = \theta_{MS} - \theta_S$.

OP, the 'second-corrected hypotenuse' ('*dvitīya-sphuṭa-karṇa*') in *yojanās* is the true distance between the Moon and the centre of the Earth. The *bhujā-phala*, ON and *koṭi-phala*, CN are given by

$$\begin{aligned} ON &= r_2 \sin(\theta_{MS} - \theta_S) \\ &= \frac{R}{2} \cos(\theta_S - \varpi) \sin(\theta_{MS} - \theta_S), \qquad (19) \\ \text{and} \quad CN &= r_2 \cos(\theta_{MS} - \varpi) \\ &= \frac{R}{2} \cos(\theta_S - \varpi) \cos(\theta_{MS} - \theta_S). \qquad (20) \end{aligned}$$

Then, the true distance is given by

$$\begin{aligned} OP = D_2 &= \sqrt{(PN)^2 + ON^2} \\ &= \sqrt{(PC + CN)^2 + ON^2} \\ &= \sqrt{(D_1 + \textit{koṭi-phala})^2 + \textit{bhujā-phala}^2} \\ &= \left[\left(D_1 + \frac{R}{2} \cos(\theta_S - \varpi) \cos(\theta_{MS} - \theta_S) \right)^2 \right. \\ &\quad \left. + \left(\frac{R}{2} \cos(\theta_S - \varpi) \sin(\theta_{MS} - \theta_S) \right)^2 \right]^{\frac{1}{2}}. \qquad (21) \end{aligned}$$

Here

$$CP = D_1 = 10K = \frac{10R^2}{\sqrt{R^2 - r_0^2 \sin^2 M - r_0 \cos M}}, \qquad (22)$$

where $M = \theta_M - \varpi$ is Moon's anomaly, and r_0 is the mean radius of the *manda*-epicycle introduced in the discussion on the equation of centre in section 4.

We now explain how *Yukti-bhāṣā* uses the above geometrical picture of lunar motion to incorporate the so called 'evection' correction to the longitude of the Moon. In Figure 7, $\theta_t = \Gamma\hat{O}P$ is the true longitude of the Moon with respect to the centre of the earth. The true position of the Moon at P can also be obtained by drawing $OP' = D_1$ in the direction of the *manda*-corrected planet with respect to O and $P'P = r_2$ from P', in the direction of the Sun.

Now

$$O\hat{P}C = P\hat{O}P' = \Gamma\hat{O}P' - \Gamma\hat{O}P = \Gamma\hat{C}P - \Gamma\hat{O}P = \theta_{MS} - \theta_t.$$

Hence

$$\sin(\theta_{MS} - \theta_t) = \sin(O\hat{P}C) = \frac{ON}{OP} = \frac{\frac{R}{2}\cos(\theta_S - \varpi)\sin(\theta_{MS} - S)}{D_2}, \quad (23)$$

where D_2 is given by (21). Hence

$$\begin{aligned} \theta_t = \theta_{MS} \quad &- \quad \sin^{-1}\left[\frac{\frac{R}{2}\cos(\theta_S - \varpi)\sin(\theta_{MS} - \theta_S)}{D_2}\right] \\ = \theta_M \quad &- \quad \sin^{-1}\left(\frac{r_0 \sin M}{R}\right) \\ &- \quad \sin^{-1}\left[\frac{\frac{R}{2}\cos(\theta_S - \varpi)\sin(\theta_{MS} - \theta_S)}{D_2}\right]. \end{aligned} \quad (24)$$

Now when $x \ll 1$, $\sin^{-1} x \approx x$ (radians) $= 3438x$ (minutes). $\frac{r_0}{R}$ for the Moon is specified to be $\frac{7}{80}$ in *Tantrasaṅgraha*. Then to the first order in $\frac{r_0}{R}$ and $\frac{R}{2D_2} \approx \frac{1}{20}$, we obtain the true longitude of the Moon to be

$$\theta_t(\text{in minutes}) = \theta_M - 300.82' \sin M - 171.9' \cos(\theta_S - \varpi)\sin(\theta_{MS} - \theta_S), \quad (25)$$

where θ_M is the Mean longitude of the Moon in minutes, the second term is the usual equation of centre and the other term is the second correction as prescribed in *Tantrasaṅgraha*. As we mentioned ealier, the second correction appears already in *Laghumānasa* [10]. There it is stated to be - $143'58'' \cos(\theta_S - \varpi)\sin(\theta_{MS} - \theta_S)$. It is clear that the lunar theory in *Tantrasaṅgraha/Yukti-bhāṣā* is very similar to that in *Laghumānasa*, though the parameter is altered.

Now, according to the modern lunar theory, there are nearly 1500 terms contributing to Moon's longitude! Most of them are negigible. Considering only the major terms, Moon's true longitude, θ_t in minutes is given by the following expression [25]:[12]

$$\begin{aligned} \theta_t(\text{in minutes}) \; &= \; \theta_M - 377.33' \sin M + 12.82' \sin(2M) \\ &- \; 76.43' \sin(2(\theta_M - \theta_S) - M) + 39.5' \sin(2(\theta_M - \theta_S)) \\ &+ \; 11.13' \sin(\theta_S - \varpi_S), \end{aligned} \quad (26)$$

[12] We have recast the formula in terms of the longitude of the apogee, instead of the perigee as in modern notation. This will only change the signs of some terms.

where ϖ_S is the longitude of Sun's apogee. In the RHS, the second and third terms represent the 'Equation of centre', and the fourth, fifth and the sixth ones are the 'Evection', 'Variation', and 'Annual inequality' repectively. Part of the second term and the fourth term can be combined using the relation:

$$\sin A + \sin B = 2\sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right).$$

Then, writing only the terms corresponding to the two inequalities, namely, 'equation of centre' and 'evection',

$$\begin{aligned}\theta_t(\text{Modern}) &= \theta_M - 300.9\sin M + 12.82\sin(2M)\\ &- 152.86\cos(\theta_S - \varpi)\sin(\theta_{MS} - \theta_S) + \text{Variation}\\ &+ \text{Annual inequality} + \ldots.. \qquad (27)\end{aligned}$$

We may now compare the expression for θ_t in *Tantrasaṅgraha* and *Yukti-bhāṣā* to the first order in $\frac{r_0}{R}$, and $\frac{R}{D_2}$ with the modern expression, and with the one which follows from the geometrical model of Ibn al-Shāṭir (outlined in Appendix 3). The coefficient of the $\sin M$ term in *Tantrasaṅgraha/Yukti-bhāṣā* formula (25) has very nearly the same value as in the modern expression (27). Ibn al-Shāṭir's value for the same is 296.05'. The evection term in (25) has the same form as in the modern expression , though with a higher coefficient of 171.9′. In fact, the *Laghumānasa* value of 143′58″ seems to be better. In Ibn al-Shāṭir's theory this has the value 162.35'. It should however be remembered that we have retained only the first order terms in the *Tantrasaṅgraha/Yukti-bhāṣā* expression, as also in Ibn al-Shāṭir's theory. There is no $\sin(2M)$ term in *Tantrasaṅgraha/Yukti-bhāṣā*. This is true even if we include the higher order terms in $\frac{r_0}{R}$. However the coefficient of this term is small even in the modern theory.

It is possible to compare the exact values of the correction to the mean longitude due to equation of centre and evection, without any approximation, at syzygy ($\theta_M - \theta_S = 0°$ or$180°$) or quadrature ($\theta_M - \theta_S = \pm 90°$), when it is also assumed that the anomaly, $M = 90°$. At syzygy, the second correction is zero in *Tantrasaṅgraha/Yukti-bhāṣā*, and when the anomaly is $90°$, $\theta_t - \theta_M = \sin^{-1}\left(\frac{7}{80}\right) =$ 301.2′, compared with the modern value of 300.9′, Ptolemy's value of 301′ and Ibn al-Shāṭir's value of 295′. At quadrature, when the anomaly is $90°$, it is easy to check that $D_1 = 10.038R$ and $D_2 = 10.05R$. Then $\theta_t - \theta_M = \sin^{-1}\left(\frac{7}{80}\right) - \sin^{-1}\left(\frac{1}{20.1}\right) =$ 472.3′, compared with the modern value of 453.8′, Ptolemy's value of 460′ and Ibn al-Shāṭir's value of 456′.

Hence, the *Tantrasaṅgraha/Yukti-bhāṣā* value for the maximum correction due equation of centre and evection is very accurate at syzygies, and reasonable at quadratures.

Variation in Earth-Moon Distance

From (22) and using the value of $\frac{r_0}{R} = \frac{7}{80}$, we find that the *manda*-hypotenuse in *yojanās*, D_1, has the values $10.9589R$, $10.0385R$ and $9.1954R$ when M is 0° (Moon at apogee), 90° and 180° (Moon at perigee), respectively.

At syzygy, $\theta_{MS} - \theta_S = 0°$ or 180°. Then from (21), $D_2 = D_1 + \frac{R}{2}\cos M$, and it can be easily seen that the true distance of the Moon from the earth, D_2, has the values $11.4589R$, $10.0385R$ and $8.6954R$ when M is 0° (apogee), 90° and 180° (perigee), respectively. Hence, at syzygy, the ratio of the maximum and minimum distances of the Moon from earth is 1.3178 according to *Tantrasaṅgraha/Yukti-bhāṣā*. This ratio is 1.1925 in Ptolemy's theory, and 1.1884 in Ibn al-Shāṭir's theory.

At quadrature, $\theta_{MS} - \theta_S = \pm 90°$. Then, $D_2 = D_1$ when $M = 0°$ or 180°, and $D_2 = \sqrt{D_1^2 + \frac{R^2}{4}}$ when $M = \pm 90°$. The true lunar distance D_2 has the values $10.9589R$, $10.0509R$ and $9.1584R$ when M is 0° (apogee), 90° and 180° (perigee), respectively. Hence, at quadrature, the ratio of the maximum and minimum distances of the Moon from earth is 1.1918 according to *Tantrasaṅgraha/Yukti-bhāṣā*. In Ptolemy's model this ratio turns out to be 1.3112 and in Ibn al-Shāṭir's model, 1.3077.[13]

Overall, the ratio of the maximum and minimum distances turns out to be 1.3178, 1.9278 and 1.3112 according to *Tantrasaṅgraha/Yukti-bhāṣā*, Ptolemy's model and Ibn al-Shāṭir's theory respectively.

Thus it is remarkable that Indian astronomers, from the time of Vaṭeśvara (c.904) and Mañjulācārya (c.932), arrived at a very accurate expression for the second lunar correction and this seems to have been based on a fairly reasonable geometrical picture of motion as revealed in the later Keraḷa works, *Tantrasaṅgraha* (c.1500) and *Yukti-bhāṣā* (c.1530).

Acknowledgements

This article grew out of my lecture in a series of seminars on History of Mathematics at the Chennai Mathematical Institute during January-February, 2008. I thank the organisers of this series as well as C.S.Seshadri, Director, CMI for arranging a very stimulating and useful set of talks. My collaborators, M.D.Srinivas, Centre for

[13] It is interesting to note that the ratio of the maximum and minimum distances at syzygy in *Tantrasaṅgraha/Yukti-bhāṣā* is close to the corresponding ratio at quadrature in Ibn al-Shāṭir's lunar theory, and viceversa. The model in *Yukti-bhāṣā* is equivalent to the introduction of an epicycle of variable radius $\frac{R}{2}\cos(\theta_S - \varpi)$ around the *Manda-spuṭa* (with respect to O) at P', in Figure 7. In this epicycle, the planet is located in the direction of the Sun. In Ibn al-Shāṭir's model, the second epicycle has a fixed radius, and the planet is located at an angle of $2(\theta_M - \theta_S)$ with respect to the apogee-line of the first epicycle, in Figure 13. These two models are associated with the two different forms of the evection term in (26) and (27).

Policy Studies, Chennai and K.Ramasubranian, IIT, Bombay have made valuable comments and helped me in preparing this article. I am greatful to them for their help. I also thank the two anonymous referees for their valuable suggestions to improve the manuscript.

References

[1] K. S. Shukla and K. V. Sarma, *Āryabhaṭīya*, Ed. with translation and notes, Indian National Science Academy, New Delhi, 1976.

[2] G. J. Toomer, *Ptolemy's Almagest*, Duckworth, London, 1984.

[3] N. M. Sverdlow and O.Neugebauer, *Mathematical Astronomy in Copernicus's De Revolutionibus*, Springer Verlag, New York, 1984.

[4] S. K. Pillai, *Tantrasaṅgraha of Nīlakaṇṭha with the Prose Commentary Laghuvivṛti by* Śaṅkara Vāriyar, Trivandrum Sanskrit Series no.188, Trivandrum, 1958; K. V. Sarma, *Tantrasaṅgraha with the Commentary Yukti-dīpikā by* Śaṅkara Vāriyar, V.V.B. Institute, Panjab University, Hoshiarpur, 1977.

[5] K. V. Sarma and V. S. Narasimhan, *Tantrasaṅgraha of Nīlakaṇṭha Somayājī*, Text (KVS) and Translation (VSN), Supplement to the Indian Journal of History of Science, INSA, New Delhi, **33**, pp.148, 1998.

[6] K. Ramasubramanian and M. S. Sriram, *Tantrasaṅgraha of Nīlakaṇṭha Somayājī*, Translation with explanatory notes, in press.

[7] K. Ramasubramanian, M. D. Srinivas and M. S. Sriram, 'Modification of the earlier Indian planetary theory by the Kerala astronomers (c.1500 AD) and the implied heliocentric picture of planetary motion', Current Science, **66**, 784, 1994.

[8] M. S. Sriram, K. Ramasubramanian and M. D. Srinivas, Eds., *500 years of Tantrasaṅgraha: A Landmark in the History of Astronomy*, Indian Institute of Advanced Study, Shimla, 2002.

[9] K. V. Sarma, K. Ramasubramanian, M. D. Srinivas and M. S. Sriram, *Gaṇita-Yukti-Bhāṣā of Jyeṣṭhadeva*, Edition of Malayalam text and English translation (KVS) and Explanatory notes (KR, MDS, MSS), Vol. I: Mathematics, Vol. II: Astronomy, Hindustan Book Agency, New Delhi, 2008; Rep. Springer Verlag, New York, 2009.

[10] Victor Roberts, 'The solar and lunar theory of Ibn ash-Shāṭir, A pre-Copernican Copernican model', Isis, **48**, pp 428–432, 1957; reprinted in E.S.Kennedy and Imad Ghanem, Eds., *The Life and Work of Ibn al-Shāṭir, An Arab Astronomer of the Fourteenth Century*, University of Aleppo, 1976.

[11] K. S. Shukla, *Laghumānasa of Mañjula*, Ed. with translation and notes, Indian National Science Academy, New Delhi, 1990.

[12] See for instance, S. N. Sen, 'Astronomy' in D. M. Bose, S. N. Sen and B. V. Subbarayappa, *A Concise History of Science in India*, Indian National Science Academy, New Delhi, 1971; S. N. Sen and K. S. Shukla, Eds. *History of astronomy in India*, Indian National Science Academy, New Delhi, 1985, and references therein; S. Balachandra Rao, *Indian Astronomy – an Introduction*, Universities Press, Hyderabad, 2000; B. V. Subbarayappa,*The Tradition of Astronomy in India, Jyotiḥśāstra*, Vol. IV, Part 4, Centre for Studies in Civilizations, New Delhi, 2008.

[13] M. S. Sriram, 'Planetary models in Indian and Greek astronomical traditions' in reference 8, p.29–58.

[14] K. V. Sarma, *A History of the Kerala School of Hindu Astronomy in perspective*, Vishveshvaranand Institude, Hoshiarpur, 1972.

[15] *Gaṇita-Yukti-bhāṣā*, p.472–483 (Translation), p.622–633 (Explanatory notes).

[16] See for instance, K.S.Shukla, *Mahābhāskarīya*, Ed. with translation and notes, p.111–119, Lucknow, 1960.

[17] p.480–484 in reference 2.

[18] *Gaṇita-Yukti-bhāṣā*, p.488–493 (Translation), p.642–648 (Explanatory notes).
[19] *Gaṇita-Yukti-bhāṣā*, p.493–495 (Translation), p.648–652 (Explanatory notes).
[20] *Gaṇita-Yukti-bhāṣā*, p.495–499 (Translation), p.653–657 (Explanatory notes).
[21] *Gaṇita-Yukti-bhāṣā*, p.496–497 (Translation).
[22] N. M. Swerdlow, 'Ptolemy's theories of the latitude of the planets in the *Almagest, Handy Tables* and *Planetary Hypotheses*' in J. Z. Buchwald and A. Franklin (eds.), *Wrong for the Right Reasons*, p.41–71, Springer, 2005.
[23] *Gaṇita-Yukti-bhāṣā*, p.584–587 (Translation), p.786–788 (Explanatory notes).
[24] Dennis W. Duke, 'The second lunar anomaly in ancient Indian astronomy', Arch. Hist. Exact. Sci. **61**, p.147–157, 2007.
[25] O. Montenbruck and T. Pfleger, *Astronomy on the Personal Computer*, Springer, Berlin, Revised edition, 2000.
[26] p.220–234 in reference 2; for a simple account see, J. L. E. Dryer, *A History of Astronomy from Thales to Kepler*, p.193–196, Second Edition, Dover, New York, 1953.

Appendix 1

Geocentric Coordinates of a Planet in Kepler's Model

The planetay models in ancient times can be appreciated better, if we understand how the geocentric coordinates of a planet is calculated in Kepler's model, which is correct in its essentials even today [13]. The three laws discovered by Kepler in the early seventeenth century form the basis of our present understanding of planetary orbits. The three laws are:

1. Each planet moves around the Sun in an ellipse, with Sun at one of the foci.
2. The areal velocity of a planet in its orbit is a constant.
3. The square of the time-period of a planet in its orbit is proportional to the cube of the semi-major axis of the ellipse in which it moves.

Kepler's laws can be derived from Newton's second law of motion and law of gravitation[14]. If we take the time period of a planet as an input, Kepler's third law does not play any role for an individual orbit. We should note that Kepler's laws are essentially kinematical laws, which do not make any reference to the concepts of 'acceleration' and 'force', as we understand them. Even then, they capture the very essence of the nature of planetary orbits and can be used to calculate the planetary coordinates, once we know the parameters of the ellipse and the initial coordinates. The planetary models in ancient astronomy are kinematical and in that sense they can be compared with Kepler's model. In the following, we shall elaborate on the computation of the geocentric longitude and latitude of a planet in Kepler's model.

[14]Conversely, Newton's inverse square law of gravitation can also be arrived at from Kepler's laws!

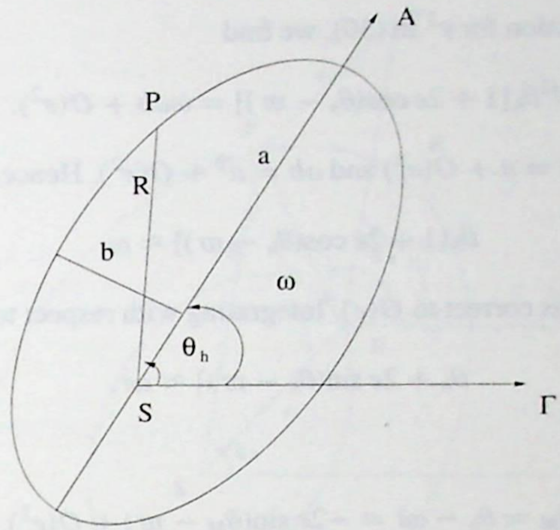


Figure 8: Elliptic orbit of a planet around the Sun.

Elliptic Orbits and Equation of Centre

The elliptic orbit of a planet (P) around the Sun (S, which is also one of the foci of the ellipse) is represented in Figure 8. a and b are the semi-major and semi-minor axes of the ellipse. Γ refers to the first point of Aries. $\varpi = \Gamma\hat{S}A$ is the longitude of the aphelion (A). $\theta_h = \Gamma\hat{S}P$ is the heliocentric longitude of the planet.

Here e is the eccentricity of the ellipse and $l = a(1 - e^2)$.

Then the equation of the ellipse (in polar coordinates, with the origin at one of the foci), may be written as

$$\frac{l}{r} = 1 - e\cos(\theta_h - \varpi). \tag{28}$$

Therefore,

$$r = l[1 + e\cos(\theta_h - \varpi)] + O(e^2), \tag{29}$$
$$r^2 = l^2[1 + 2e\cos(\theta_h - \varpi)] + O(e^2). \tag{30}$$

Now the areal velocity of the planet at any instant is $\frac{1}{2}r^2\dot{\theta}_h$. According to Kepler's second law, this is a constant. As the area of an ellipse is πab, the areal velocity can also be written as $\frac{\pi ab}{T} = \frac{\omega ab}{2}$, where T is the time period and $\omega = \frac{2\pi}{T}$ is the mean angular velocity of the planet. Hence,

$$r^2\dot{\theta}_h = \omega ab. \tag{31}$$

Using the expression for r^2 in (30), we find

$$l^2\dot{\theta}_h[1 + 2e\cos(\theta_h - \varpi)] = \omega ab + O(e^2). \quad (32)$$

Now $l = a\,(1 - e^2) = a + O(e^2)$ and $ab = a^2 + O(e^2)$. Hence,

$$\dot{\theta}_h[1 + 2e\cos(\theta_h - \varpi)] \approx \omega, \quad (33)$$

where the equation is correct to $O(e)$. Integrating with respect to time, we obtain

$$\theta_h + 2e\sin(\theta_h - \varpi)] \approx \omega t, \quad (34)$$

or again,

$$\theta_h - \theta_M = \theta_h - \omega t = -2e\sin(\theta_M - \varpi) + O(e^2). \quad (35)$$

Here $\theta_M = \omega t$ is the mean longitude which increases linearly with time, t. $\omega t - \varpi$, the difference between the longitudes of the mean planet and the apogee is the 'anomaly'. Eq.(35) gives the equation of centre which is the difference between the true heliocentric longitude θ_h and the mean longitude θ_M, correct to $O(e)$, in terms of the anomaly. Clearly, the equation of centre arises due to the eccentricity of the orbit.

Geocentric Longitude of an Exterior Planet

The orbits of all the planets are inclined at small angles to the plane of Earth's orbit around the Sun, known as ecliptic. We will ignore these inclinations and assume that all the planetary orbits lie on the plane of the ecliptic while calculating the planetary longitudes, as the corrections introduced by them (inclinations) are known to be small. We will consider the longitude of an exterior planet like Mars, Jupiter or Saturn and an interior planet like Mercury or Venus, separately.

The elliptic orbit of an exterior planet (P) and that of the Earth (E) around the Sun (S) are shown in Figure 9. Here, $\theta_h = \Gamma\hat{S}P$ is the true heliocentric longitude of the planet. $\theta_S = \Gamma\hat{E}S$ and $\theta_g = \Gamma\hat{E}P$ are the true geocentric longitudes of the Sun and the planet respectively, while r and R are the distances of the Earth and the planet from the Sun, which vary along their orbits. To facilitate comparison with the Indian models, $EP' = R$ is drawn parallel to SP. Then $P'P$ is parallel to ES and $P'P = r$. We have already described how θ_h is computed using the expression for the equation of centre. From this, the true geocentric longitude, θ_g has to be computed.

Now

$$E\hat{P}S = P\hat{E}P' = \theta_g - \theta_h, \quad (36)$$

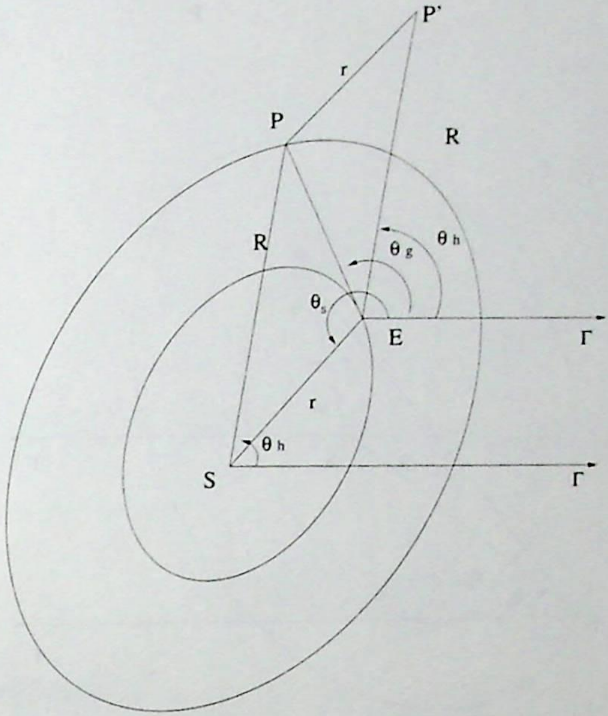


Figure 9: Heliocentric and geocentric longitudes of an exterior planet in Kepler's model.

and

$$E\hat{S}P = 180^\circ - (\theta_S - \theta_h). \tag{37}$$

$$EP^2 = R^2 + r^2 - 2rR\cos[180^\circ - (\theta_S - \theta_h)], \tag{38}$$

or,

$$EP = [(R + r\cos(\theta_S - \theta_h))^2 + r^2 \sin^2(\theta_S - \theta_h)]^{1/2} \tag{39}$$

Also,

$$\frac{\sin(E\hat{P}S)}{ES} = \frac{\sin(E\hat{S}P)}{EP}. \tag{40}$$

Using (36)–(40),

$$\sin(\theta_g - \theta_h) = \frac{r\sin(\theta_S - \theta_h)}{[(R + r\cos(\theta_S - \theta_h))^2 + r^2\sin^2(\theta_S - \theta_h)]^{1/2}}. \tag{41}$$

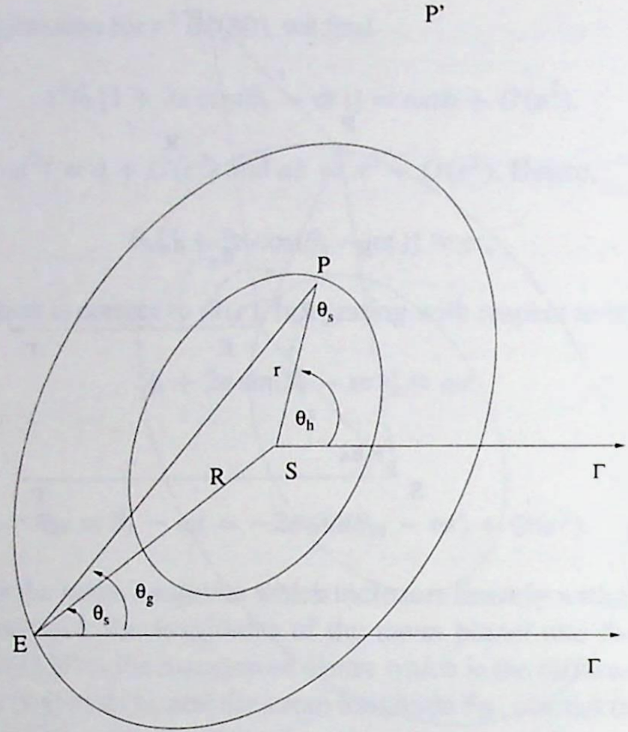


Figure 10: Heliocentric and geocentric longitudes of an interior planet in Kepler's model.

Here $(\theta_S - \theta_h)$, the difference between the longitudes of the Sun and the heliocentric planet, is the 'solar anomaly'. Thus, (41) gives $\theta_g - \theta_h$ in terms of the solar anomaly. Adding this to θ_h, we get the true geocentric longitude, θ_g of the planet.

Geocentric Longitude of an Interior Planet

The elliptic orbit of an interior planet (P) and that of the Earth (E) around the Sun are shown in Figure 10. Here, $\theta_h = \Gamma\hat{S}P$ is the true heliocentric longitude of the planet, which can be computed from the mean heliocentric longitude and the equation of centre. $\theta_S = \Gamma\hat{E}S$ and $\theta_g = \Gamma\hat{E}P$ are the true geocentric longitudes of the Sun and the planet respectively. Here r and R are the variable distances of the planet and the Earth from the Sun, respectively.

Now

$$S\hat{E}P = \theta_g - \theta_S. \tag{42}$$

It can be easily seen that

$$E\hat{S}P = 180° - (\theta_h - \theta_S), \tag{43}$$

$$EP = [(R + r\cos(\theta_h - \theta_S))^2 + r^2\sin^2(\theta_h - \theta_S)]^{1/2}. \tag{44}$$

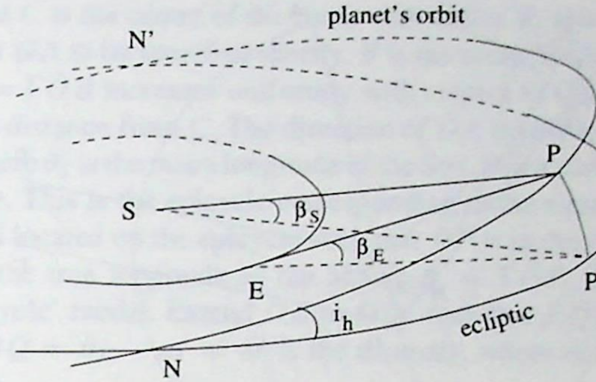


Figure 11: Heliocentric and geocentric latitudes of a planet in Kepler's model.

Also,

$$\frac{\sin(S\hat{E}P)}{SP} = \frac{\sin(E\hat{S}P)}{EP}. \tag{45}$$

Using (42)–(45) we get,

$$\sin(\theta_g - \theta_S) = \frac{r\sin(\theta_h - \theta_S)}{[(R + r\cos(\theta_h - \theta_S))^2 + r^2\sin^2(\theta_h - \theta_S)]^{1/2}}. \tag{46}$$

The difference $(\theta_g - \theta_S)$ is determined from this equation. Adding this to θ_S, we get the true geocentric longitude, θ_g of the planet. Note that the true longitude of the Sun, θ_S, and the true heliocentric longitude of the planet, θ_h (obtained by adding the equation of centre for the planet to the mean heliocentric longitude longitude) should be obtained first. Then the true geocentric longitude of an interior planet can be obtained using (46).

Heliocentric and Geocentric Latitudes of a Planet

In Figure 11, the orbit of the planet P is inclined at an angle i_h to the ecliptic. N and N' are the nodes. PP' is the circular arc perpendiular to the ecliptic. Then the heliocentric latitude β_S is given by

$$\beta_S = PSP' = \frac{PP'}{SP}. \tag{47}$$

If λ_P and λ_N are the heliocentric longitudes of the planet and the node, it can be easily seen that

$$\sin\beta_S = \sin i_h \, \sin(\lambda_P - \lambda_N), \tag{48}$$

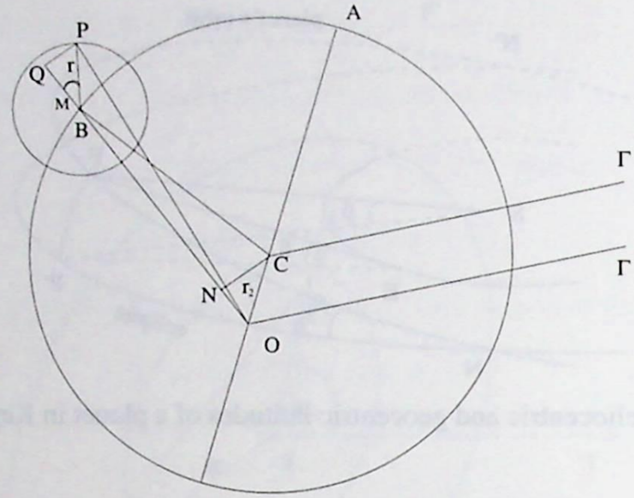


Figure 12: Ptolemy's model of lunar motion.

or,

$$\beta_S \approx i_h \ \sin(\lambda_P - \lambda_N), \tag{49}$$

as i_h and β_S are small. Note that λ_P stands for the true heliocentric longitude of the planet. We have also shown Earth's orbit in the figure. The geocentric latitude β_E is given by

$$\beta_E = PEP' = \frac{PP'}{EP}. \tag{50}$$

From (47)–(50), we find that

$$\begin{aligned}\beta_E &= \beta_S \frac{SP}{EP} \\ &= \frac{i_h \ SP \ \sin(\lambda_P - \lambda_N)}{EP},\end{aligned} \tag{51}$$

where EP, the true distance of the planet from the Earth, can be found from (39) or (44).

Appendix 2

Ptolemy's Lunar Theory

Figure 12 depicts the geometrical model of Ptolemy which incorporates both the corrections, namely, 'equation of centre' and 'evection' [26]. Here O is the centre

of the earth and C is the centre of the circle with radius R, at a distance r_2 from O in the direction OA to be specified shortly. B is the mean Moon, whose geocentric longitude $\theta_M = \Gamma\hat{O}B$ increases uniformly with respect to O and not C, though it is at a constant distance from C. The direction of OA is chosen such that $A\hat{O}B = 2(\theta_M - \theta_S)$, where θ_S is the mean longitude of the Sun. Draw a circle of radius r with B as the centre. This is the epicycle corresponding to the equation of centre. The true Moon P is located on the epicycle such that BP is in the direction of Moon's apogee. Then the true longitude of the Moon, $\theta_t = \Gamma\hat{O}P$. This is a combined 'eccentric-epicycle' model. Extend OB to OQ, such that PQ is perpendicular to OQ. Then $P\hat{B}Q = \theta_M - \varpi = M$ is the anomaly, where ϖ is the longitude of Moon's apogee.

The difference between the true and mean longitudes of the Moon, $\theta_M - \theta_t = \Gamma\hat{O}B - \Gamma\hat{O}P = P\hat{O}B$, satisfies the relation:

$$OP\sin(\theta_M - \theta_t) = PQ = r\sin M. \tag{52}$$

$$\text{Now } OP = \sqrt{(OB + r\cos M)^2 + r^2\sin^2 M}. \tag{53}$$

Here OB is to be determined. Draw CN perpendicular to OB. As $C\hat{O}N = 2(\theta_M - \theta_S)$, we have $CN = r_2\sin 2(\theta_M - \theta_S)$ and $ON = r_2\cos 2(\theta_M - \theta_S)$. Noting that $NB = \sqrt{CB^2 - CN^2}$, with $CB = R$, we have

$$OB = NB + ON = \sqrt{R^2 - r_2^2\sin^2(2(\theta_M - \theta_S))} + r_2\cos(2(\theta_M - \theta_S)). \tag{54}$$

Then the correction to the mean longitude, $\theta_t - \theta_M$ is to be determined from

$$\sin(\theta_M - \theta_t) = \frac{r\sin M}{\sqrt{(OB + r\cos M)^2 + r^2\sin^2 M}}, \tag{55}$$

where OB is given by (54).

The above equation (55) incorporates both the 'equation of centre' and 'evection' terms. The expression for $\theta_t - \theta_M$ does not reduce to the modern form (27) for the sum of these terms, in any reasonable approximation. However at syzygy, the correction has the same form as the $\sin M$ term in the equation of centre with $OB = R + r_2$ replacing R. The second correction would be zero at quadrature also, when $M = 0°$ or $180°$ (at perigee or apogee of the epicycle), just as in the modern theory. The second correction has the maximum value when $M = \pm 90°$.

Ptolemy gives the maximum value of the total correction to be 301' at syzygy, compared to the modern value of 300.9'. At quadrature the maximum is given to be 460' compared to the modern value of 453.76'. At other values of $\theta_M - \theta_S$ and anomaly, Ptolemy's theory would not be impressive, as far as the correction to the mean longitude is concerned.

The values of $\frac{r}{R}$ and $\frac{r_2}{R}$ can be calculated from the maximum corrections at the syzygy and quadrature. From (54) and (55), it can be easily checked that

$$\text{at syzygy, } tan(\theta_M - \theta_t)|_{maximum} = \frac{r}{R + r_2} \quad (56)$$

$$\text{and at quadrature, } tan(\theta_M - \theta_t)|_{maximum} = \frac{r}{R - r_2} \quad (57)$$

From the given values of the maximum values, we find that $\frac{r}{R} = 0.1063$ and $\frac{r_2}{R} = 0.2106$ in Ptolemy's model.

Variation in Earth-Moon Distance

Now at syzygy, $OB = R + r_2$. Then $OP = R + r_2 + r$, when $M = 0°$ (apogee of the epicycle), and $OP = R + r_2 - r$, when $M = 180°$ (perigee of the epicycle). Similarly, at quadrature, $OB = R - r_2$. Then $OP = R - r_2 + r$ when $M = 0°$ (apogee of the epicycle), and $OP = R - r_2 - r$ when $M = 180°$ (perigee of the epicycle). Using the values of $\frac{r}{R}$ and $\frac{r_2}{R}$ above, we find that the ratio of the maximum and minimum distances of the Moon from earth is 1.1925 at syzygy, 1.3112 at quadrature, and 1.9278 overall. This is of course absurd. Hence, Ptolemy's model can be hardly taken to be a serious physical model. This is one of the reasons for the search for an alternate lunar theory in the later Islamic tradition of astronomy, which otherwise followed Ptolemaic framework.

Appendix 3

The Lunar Theory of Ibn al-Shāṭir

Ibn al-Shāṭir (1304-1375/6 CE) developed a lunar theory in his *Kitāb Nihāyat as-Sūl fi Taṣḥīḥ al-Uṣūl* , where he dispenses with the eccentric deferent in Ptolemy's theory and introduces instead a second epicycle [10]. His model is illustrated in Figure 13. Here O is the centre of the earth. The mean Moon moves on a deferent circle of radius R. Consider the case when the mean Moon, mean Sun and Moon's apogee are in the same direction. Here the mean Moon is at A ($OA = R$). Draw a circle of radius a around A, which is the epicycle corresponding to the 'first inequality'. A_1 is apogee of the epicycle. Draw a second epicycle of radius b around A_1. Then the true Moon is at B which is the perigee of the second epicycle.

Consider the case when the mean Moon is at P_0, and the anomaly, $A\hat{O}P_0 = M$. Draw $P_0A_2 = a$ in the direction of Moon's apogee. Draw $A_2P = b$, such that $P_0\hat{A}_2P = 2(\theta_M - \theta_S)$, that is, the radius of the second epicycle rotates at twice the rate of elongation of the mean Moon from the mean Sun, eastwards (counterclockwise) from the direction of the apogee. Then the true Moon is located at P. Draw TQ

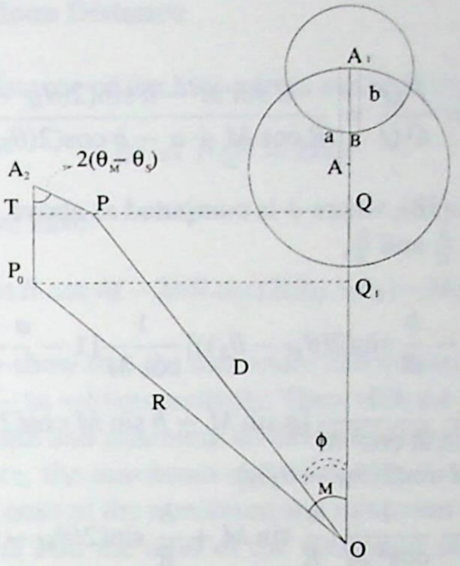


Figure 13: Ibn al-Shāṭir's model for the two corrections to the Moon.

through P and P_0Q_1 perpendicular to OA and P_0A_2. Let ϕ be the angle between the directions of the true Moon (OP) and the apogee (OP_1). Let θ_t be Moon's true longitude, and $\delta\theta = \theta_t - \theta_M$ be the total inequality. Then the 'total inequality' $\delta\theta$ can be expressed as

$$\delta\theta = \theta_t - \theta_M = -P\hat{O}P_0 = \phi - M \tag{58}$$

We derive the expression for the total inequality in the above geometrical model, and simplify it when only terms which are first order in $\frac{a}{R}$ and $\frac{b}{R}$ are retained. This will facilitate comparison with the modern expression for the lunar inequalities given by (26) and (27). Now,

$$\begin{aligned} TQ = P_0Q_1 &= R\sin M, \\ TP = A_2P\sin(2(\theta_M - \theta_s)) &= b\sin(2(\theta_M - \theta_S)). \end{aligned} \tag{59}$$

Hence,

$$\begin{aligned} PQ = TQ - TP &= R\sin M - b\sin(2(\theta_M - \theta_S)) \qquad (60) \\ A_2T = A_2P\cos(2(\theta_M - \theta_s)) &= b\cos(2(\theta_M - \theta_S)), \\ P_0A_2 &= a, \\ QQ_1 = P_0T = P_0A_2 - A_2T &= a - b\cos(2(\theta_M - \theta_S)), \\ OQ_1 &= R\cos M, \\ OQ = OQ_1 + QQ_1 &= R\cos M + a - b\cos(2(\theta_M - \theta_S)). \qquad (61) \end{aligned}$$

Hence

$$tan(\phi) = \frac{PQ}{OQ} = \frac{R \sin M - b \sin(2(\theta_M - \theta_S))}{R \cos M + a - b \cos(2(\theta_M - \theta_S))}. \quad (62)$$

Then $\delta\theta$ is found from (58), where ϕ is computed as above.

To the first order in $\frac{a}{R}$ and $\frac{b}{R}$,

$$\begin{aligned}\tan(\phi) &\approx [\sin M - \frac{b}{R}\sin(2(\theta_M - \theta_S))]\frac{1}{\cos M}[1 - \frac{a - b\cos(2(\theta_M - \theta_S))}{R \cos M}] \\ &\approx \tan M - \frac{1}{R\cos^2 M}[a \sin M + b \sin M \cos(2(\theta_M - \theta_S)) \\ &- b \cos M \sin(2(\theta_M - \theta_S))] \\ &= \tan M - \frac{1}{\cos^2 M}[\frac{a}{R}\sin M + \frac{b}{R}\sin(2(\theta_M - \theta_S) - M)]. \quad (63)\end{aligned}$$

When the total correction is small,

$$\phi - M \approx \tan(\phi - M) \approx (\tan\phi - \tan M)\cos^2 M.$$

Then

$$\begin{aligned}\theta_t &\approx \theta_M - [\frac{a}{R}\sin M + \frac{b}{R}\sin(2(\theta_M - \theta_S) - M)] \\ &\approx \theta_M - [\frac{a-b}{R}\sin M + \frac{2b}{R}\sin(\theta_M - \theta_S)\cos(\theta_S - \varpi)], \quad (64)\end{aligned}$$

where ϖ is the longitude of Moon's apogee ($M = \theta_M - \varpi$).

Comparing the above with the modern expression (27) for the lunar inequalities, we notice the absence of the $\sin(2M)$ term in Ibn al-Shāṭir's theory. $\frac{a}{R}$ and $\frac{b}{R}$ are given to be $\frac{6+35/60}{60}$ and $\frac{1+25/60}{60}$ in his model. Expressing the result in minutes,

$$\theta_t \approx \theta_M - 296.05' \sin M - 162.35' \sin(\theta_M - \theta_S)\cos(\theta_S - \varpi). \quad (65)$$

This is to be compared with the values 300.9 and 152.864 for the coefficients of the $\sin M$ and the other term in the modern expression (27).

It is possible to compare the exact values of the maximum correction to the mean longitude due to equation of centre and evection, without any approximation, at syzygy ($\theta_M - \theta_S = 0°$ or180°) or quadrature ($\theta_M - \theta_S = \pm 90°$), when the anomaly, $M = 90°$. With the given values of a and b, the maximum correction ($\theta_M - \theta_t = M - \phi$) has the values 295′ at syzygy and 456′ at quadrature, compared to the modern values 300.9′ and 453.8′ respectively.

Variation of Earth-Moon Distance

From Figure 13, the distance of the Moon from earth, D is given by

$$D^2 = PQ^2 + OQ^2. \tag{66}$$

Using (61) and (62), we have

$$D^2 = R^2+a^2+b^2+2aR\cos M-2bR\cos(2(\theta_M-\theta_S)-M)-2ab\cos(2(\theta_M-\theta_S)). \tag{67}$$

From this, it is easy to show that the maximum and minimum distances at syzygy are $R+(a-b)$ and $R-(a-b)$ respectively. Then, with the given values of a and b, the ratio of the maximum and minimum distances at syzygy is found to be 1.1884. Similarly, at quadrature, the maximum and minimum distances are $R+(a+b)$ and $R-(a+b)$. The ratio of the maximum and minimum distances at quadrature is then, 1.3077. This is also the ratio of the maximum and minimum distances, overall. Clearly, Ibn al-Shāṭir does manage to modify the Ptolemaic model so as to get rid of the totally unreasonable variation in the Earth-Moon distance which was a longstanding problem in the development of lunar theory in the Greek and Islamic traditions which followed the Ptolemaic framework.

M. S. Sriram
Department of Theoretical Physics
University of Madras
Guindy Campus, Chennai 600 025, India.
E-mail: Sriram.physics@gmail.com

SLS

This volume is the outcome of a Seminar on History of Mathematics held at the Chennai Mathematical Institute during January-February, 2008, and contains articles based on talks by distinguished scholars both from the West and from India.

The topics covered include: Geometry in the *Śulvasūtras*; Origins of zero (which can be traced to ideas of *lopa* in Pāṇini's grammar); Combinatorial methods in Indian Music (which were developed in the context of prosody and subsequently applied to the study of tonal and rhythmic patterns in music); Cross-cultural view of the development of negative numbers (from Brahmagupta (c.628 CE) to John Wallis (1685 CE)); Kuṭṭaka, Bhāvanā and Cakravāla (the techniques developed by Indian mathematicians for the solution of indeterminate equations); Development of calculus in India (covering the millennium long history of discoveries culminating in the work of the Kerala School giving a complete analysis of the basic calculus of polynomial and trigonometric functions); Recursive methods in Indian Mathematics (going back to Pāṇini's grammar and culminating in the recursive proofs found in the Malayalam Text *Yuktibhāṣā* (1530 CE)); and Planetary and Lunar models developed by the Kerala School of Astronomy. These articles in their totality cover a substantial portion of the History of Indian Mathematics and Astronomy.

This book would serve the dual purpose of bringing to the international community a better perspective of the mathematical heritage of India, and at the same time convey the message that much work yet remains to be done, namely the study of many unexplored manuscripts still available in libraries in India and abroad.

ISBN 978-93-80250-06-9

9 789380 250069

www.hindbook.com